

UNIVERSITETET I OSLO
Institutt for informatikk

Masteroppgave

Lydfokusering og
kildeseparasjon i et
lokalt ekkofylt miljø

Bjørn Cato Syversrud

1. februar 2007



Innhold

| | |
|---|------------|
| Forord | v |
| Sammendrag | vii |
| 1 Introduksjon | 1 |
| 1.1 Relatert arbeid | 5 |
| 1.1.1 Forskinkelse-og-sum-stråleformer | 5 |
| 1.1.2 Blind Kildeseparasjon (BSS) | 6 |
| 1.1.3 Lydfokusering med tidsreversering | 7 |
| 1.1.4 Lydfokusering med inversfiltrering | 9 |
| 2 Teori | 11 |
| 2.1 Minste kvadraters metode | 11 |
| 2.1.1 Kolonnerom og rang | 12 |
| 2.2 Singulærverdidekomposisjon | 13 |
| 2.3 SVD og minste kvadraters metode | 13 |
| 2.4 Litt om hvordan lyd oppfører seg i rom | 14 |
| 2.5 Systemer og systemidentifikasjon | 14 |
| 2.6 Menneskets hørsel som et system | 17 |
| 2.6.1 Tidsforsinkelse og amplitudeforskjell til en innkom- mende bølge | 17 |
| 2.6.2 Hode-relaterte systemfunksjoner | 18 |
| 2.7 Inversfiltrering/dekonvolusjon | 19 |
| 2.8 Anvendelse av inversfiltrering | 21 |
| 2.9 Utvidet inversfiltrering | 21 |
| 2.9.1 Undertrykkelsesfiltre | 24 |
| 2.9.2 Kombinert inversfiltrering og undertrykkelse | 25 |
| 2.10 Kildeseparasjon og lydfokusering | 27 |
| 2.10.1 Kildeseparasjon | 27 |
| 2.10.2 Lydfokusering | 29 |
| 2.11 Separasjonskriterier | 31 |
| 2.11.1 Kriterier for å måle om hvorvidt lydfokuseringen er vellykket | 31 |

| | | |
|----------|--|-----------|
| 2.11.2 | Separasjonsmål for kilde separasjon | 33 |
| 2.12 | Impulsresponsen i et rikt miljø | 34 |
| 2.13 | Dårlig kondisjonerte likningsystemer | 34 |
| 3 | En simuleringsmodell | 37 |
| 3.1 | Visualisering av fokuseringen | 39 |
| 3.2 | Faktisk fokusering og utilsiktet undertrykking | 41 |
| 3.3 | Romlig glatting av impulsresponsene | 41 |
| 4 | Simuleringsresultater | 47 |
| 4.1 | Faktorer som undersøkes | 47 |
| 4.2 | Mengden glatting og innvirkning på fokuseringsegenskapene til modellen | 48 |
| 4.2.1 | Mengden glatting og innvirkning på den estimerte responsen | 53 |
| 4.2.2 | Mengden glatting og likhet mellom impulsresponsen | 55 |
| 4.2.3 | Hvordan mengden glatting påvirker den globale fokuseringen | 55 |
| 4.3 | Lengden på impulsresponsene og innvirkningen på fokuseringsegenskapene | 57 |
| 4.3.1 | Gjennomsnittlige relative energinivåer, relative maksimale utslag og impulsresponsen med forskjellig mengde rikdom. | 59 |
| 4.3.2 | Impulsresponslengder og globale fokuseringsegenskaper ved modellen | 60 |
| 4.4 | Avveining mellom faktisk fokusering og utilsiktet undertrykking | 66 |
| 4.4.1 | Forholdet mellom antall lydkilder og antall undertrykkingspunkter og innvirkning på faktisk fokusering og utilsiktet undertrykking | 66 |
| 4.4.2 | Lengden på de inverse filterne og påvirkning på de globale fokuseringsegenskapene til modellen. | 67 |
| 4.5 | Plasseringen av undertrykkingspunktene og innvirkning på utilsiktet undertrykking | 68 |
| 4.6 | Oppsummering | 71 |
| 5 | Eksperimenter | 73 |
| 5.1 | Undersøkelse av lydfokusering med 2 høyttalere i rom | 73 |
| 5.1.1 | Oppsettet av systemet | 74 |
| 5.1.2 | Måling av impulsresponsen | 74 |
| 5.1.3 | Simuleringer med impulsresponsen målt i rom: Sammenlikning av fokusering med impulsresponsen målt med og uten reflektorer | 76 |
| 5.1.4 | Robusthet/Støy-analyse | 77 |

| | | |
|----------|---|------------|
| 5.1.5 | Signal-støy-forhold | 79 |
| 5.1.6 | Tilpasningsfeil og relative energinivåer ved løsningen av likningsystemet med støy. | 80 |
| 5.2 | Kildeseparasjon: Undersøkelse av separasjonsegenskaper med hode-relaterte-impulsresponser | 81 |
| 5.2.1 | Et sett med hode-relaterte impulsresponser | 82 |
| 5.2.2 | Simuleringer med hode-relaterte impulsresponser . . | 82 |
| 6 | Konklusjon og diskusjon | 89 |
| 6.1 | Videre forskning | 92 |
| 6.2 | Applikasjoner | 92 |
| A | Impulsreponsmålinger | 94 |
| A.1 | Impulsresponsmålinger med sinus-sveip | 95 |
| B | NORSIG 2006 artikkelbidrag | 97 |
| C | Program-beskrivelser | 102 |
| | Bibliography | 105 |
| | List of Figures | 109 |
| | List of Tables | 113 |

Forord

Denne oppgaven er innlevert til Institutt for Informatikk ved Universitetet i Oslo som en del av min Mastergrad (M.Sc.)

Jeg vil benytte denne anledningen til takke min veileder Post. Doc. Tobias Dahl for mange kreative innspill og givende faglige diskusjoner underveis i arbeidet med denne oppgaven. Jeg retter også en takk til Professor Sverre Holm for samarbeidet med artikkelen "Source Enhancement and Suppression with binaural Head Related Impulse Response Functions", samt venner og familie for god støtte.

En spesiell takk går til min samboer Trine Skipping som har vist en stor tålmodighet og forståelse for at dette arbeidet til tider har vært altoppslukende. Din støtte har vært uvurderlig!

Bjørn Cato Syversrud
1. februar 2007.

Sammendrag

I et rom hvor mange lydkilder er tilstede samtidig har mennesker en evne til å rette oppmerksomheten mot en bestemt kilde, samtidig som de øvrige kildene undertrykkes. Mennesker klarer altså å skille kilder fra hverandre med stor nøyaktighet. Det er derfor interessant å se om de hode-relaterte impulsresponsene har en innvirkning på denne separasjonsegenskapen. Et innledende arbeid ble utført og resultatene ble publisert i en artikkel på NORSIG 2006, 7th NORDIC SIGNAL PROCESSING SYMPOSIUM.

I denne oppgaven, tas det på bakgrunn av dette, utgangspunkt i impulsresponser tilhørende et "lokalt" miljø og dette benyttes for å fokusere lyd. En antagelse om at lengre og mer innholdsrike impulsresponser vil bidra til at fokuseringen bedres, legges til grunn.

Med en modell implementert i MatlabTM undersøkes denne antagelsen ved simulering av lydfokusering. Disse undersøkelsene fortsettes på ekte, målte data, som består av impulsresponser målt i et rom, samt hode-relaterte impulsresponser. Med de hode-relaterte impulsresponsene betraktes både lydfokusering og kildeseparasjon. Resultatene fra simuleringer med modellen og simuleringer med ekte data sammenliknes.

Simuleringer med modellen viser at økt rikdom i miljøet hvor fokuseringen skal utføres, øker fokuseringsegenskapene til systemet. Økt rikdom i impulsresponsene gir bedre kontroll over lydfeltet som helhet. Det vises også at ulikhet mellom impulsresponser i rom også har en innvirkning på hvor god fokusering som kan oppnås.

Ved målinger av impulsresponser i rom er det gjort forsøk på å anskaffe impulsresponser med et rikt innhold. Dette er gjort ved å sette reflektorer foran høyttalerne. Ved sammenlikning med impulsresponser målt uten reflektorer gis det implikasjoner på at reflektorene gir ønsket effekt.

Til slutt utføres det simuleringer hvor kildeseparasjon og lydfokusering med hode-relaterte impulsresponser betraktes. Ved noen plasseringer av kilder oppnås det god separasjon, ved andre plasseringer oppnås det ingen

separasjon. Satt i sammenheng med simuleringer med modellen vises det at de hode-relaterte impulsresponsene ikke er innholdsrike nok til kunne oppnå tilstrekkelig fokusering.

Kapittel 1

Introduksjon

Utbredelse av lydbølger i akustiske omgivelser har lenge vært et interessant tema for mange forskere. Å ha kjennskap til hvordan lydbølger forplanter seg er viktig innenfor seismisk signalbehandling, ultralyd og i applikasjoner for hørbar lyd. I det sistnevnte er problemstillinger som inversfiltrering av rom-impulsresponser og blind kilde-separasjon gitt veldig mye oppmerksomhet, og det finnes utallige artikler om emnene [31], [21], [20], [25], [30].

Det å fokusere akustiske signaler, eller å kunne kontrollere et lydfelt kan i en vid forstand relateres til en rekke applikasjoner innen signalbehandling. Resultater fra forskning innen akustikk og ultralyd har vist at man kan oppnå en forbedring i fokuseringen, om miljøet fokuseringen utføres i, er uordnet. Dette har vist seg også å kunne utnyttes innenfor telekommunikasjon [28], og ved fokusering av lyd i rom [36], [37]. Lydbølgeutbredelse i uordnede medier vil resultere i mere ekkoer. Det har vist seg at disse ekkoene kan brukes konstruktivt.

Å kunne separere ut en signalkilde i et miljø hvor flere er tilstede samtidig er noe mennesker gjør med stor nøyaktighet, vi kan rette oppmerksomheten mot en bestemt kilde. Å kunne separere kilder fra hverandre er relatert til fokusering av lyd.

Hode-relaterte impulsresponser er filtre som beskriver hvordan lyden fra et punkt rundt hode til en person filtreres, før lyden når trommehinnene. Denne filtreringen avhenger av formen på hode, skuldre, bryst og det ytre øret, og av hvilken retning kilden kommer fra, og utgjør det som kalles hode-relaterte impulsresponser [12]. Hode-relaterte impulsresponser undersøkes for å få kjennskap til de ytre faktorene som mennesk-

ser lytter gjennom[2]. De brukes også til å syntetisere realistiske virtuelle lydlandskaper[12]. Er disse retningsbestemte filterne med på hjelpe en lytter til å separere ut en ønsket kilde ut blant en miks av flere? Kan retningsbestemte impulsresponser brukes til å fokusere lyd?

Problemstilling

Denne masteroppgaven er inspirert av menneskers evne til å separere ut en kilde blant en miks av flere. Et innledende spørsmål var om man kunne få innsikt i dette ved hjelp av hode-relaterte impulsresponser og en matematisk modell for kilde-separasjon. En antagelse om at de hode-relaterte impulsresponsene ville være av en slik form at de kunne hjelpe en lytter til å separere kilder fra hverandre, ble lagt til grunn. Denne antagelsen blir undersøkt videre i kontekst av lydfokusering. Med dette som utgangspunkt var det noen interessante spørsmål som oppsto:

1. Finnes det en parallell mellom det ytre ørets tilsynelatende kaotiske form, og evnen til å separere lydkilder?
2. Vil innholdsrike impulsresponser gi bedre separasjon?
3. Vil formen på hode, skuldre, bryst og det ytre øret gi opphav til lengre og mer innholdsrike impulsresponser?
4. Vil lengre og mer innholdsrike impulsresponser gi bedre fokusering?
5. Vil ulikhet i tidsdomene-representasjonen mellom impulsresponsene i rom ha en innvirkning på fokuseringen?
6. Vil man på bakgrunn av dette kunne redusere antall lydkilder i fokuseringsproblemet når det skal anvendes i applikasjoner for hørbar lyd?

Dette arbeidet er et forsøk på å besvare disse spørsmålene. Oppgaven begrenses til å betrakte et "lokalt område" der separasjonen/fokuseringen skal utføres, som det lokale området rundt hodet til et menneske.

Oppgavens omfang

Denne masteroppgaven omhandler primært fokusering av lyd, hvor jeg legger til grunn antagelsen om at impulsresponser tilhørende et rikt miljø bedrer fokuseringen.

Tidlig i dette arbeidet ble det gjort undersøkelser hvor kilde-separasjon med hode-relaterte impulsresponser ble undersøkt. Disse undersøkelsene viste

lovende resultater. Arbeidet endte i en artikkel skrevet av veilederen for denne oppgaven, Post. Doc. Tobias Dahl, Professor Sverre Holm (begge ved forskningsgruppen for digital signalbehandling og bildeanalyse), og meg selv. Artikkelen ble publisert på NORSIG 2006, 7th NORDIC SIGNAL PROCESSING SYMPOSIUM[6]. Artikkelen er vedlagt i tillegg B, og var en del av denne masteroppgaven.

Simuleringsresultatene presentert i artikkelen ga grunnlag for videre undersøkelser, som ble utført med en modell implementert i Matlab™. Disse simuleringene ble gjort med tanke på lydfokusering, og forsøkte å belyse noen av spørsmålene listet opp i problemstillingen. Lydfokusering undersøkes videre med ekte, målte data. De målte dataene er impulsresponser målt gjennom et sett med reflektorer i et rom. Resultatene fra disse simuleringene settes i sammenheng med resultatene fra simuleringene med modellen.

Til slutt betraktes lydfokusering og kildeseparasjon ved simuleringer med hode-relaterte impulsresponser. Disse avsluttende simuleringene settes også i sammenheng med resultatene fra simuleringene med modellen.

| Liste over symboler |
|--|
| A - Matrise x - Vektor $h_{i,j}(n)/h_j(\phi, \theta, n)$ - Impulsrespons $g_j(n)$ - Inverst filter $\mathbf{H}_{i,j}/\mathbf{H}_j(\phi, \theta)$ - Konvolusjonsmatrise H - Sammensatt konvolusjonsmatrise d - Vektor med de diskrete verdiene til en delta funksjon av endelig lengde 0 - Null-vektoren $\underline{\mathbf{d}} = [\mathbf{d}^T \mathbf{0}^T \dots \mathbf{0}^T]^T$ g - Vektor som inneholder filter koeffisientene til M inverse filtere M - Antall lydkilder ved lydfokusering. Antall sensorer ved kildeseparasjon L - Antall kontrollpunkter ved lydfokusering. Antall kilder ved kildeseparasjon ϵ - Tilpasningsfeil ved minste kvadraters tilpasning, angis i desibel $f_2(.,.)$ - Et mål på den faktiske fokuseringen mellom fokuspunkt og ett undertrykkelsespunkt. Angis i desibel f_2^- - Et mål på utilsiktet undertrykking. Angis i desibel SIR_{ut} - Et mål på hvor god kildeseparasjons som oppnås. |

Oppgavens struktur

Oppgaven er strukturert på følgende måte:

Kapittel 2 er et tilbakeblikk på teorien som legges til grunn i denne oppgaven.

I **Kapittel 3** defineres en modell som brukes til simulasjoner.

Kapittel 4 inneholder resultatene fra simuleringene.

I **Kapittel 5** fortsettes undersøkelsene på ekte, målte data, hvor lydfokusering betraktes først. Til slutt betraktes kildeseparasjon og lydfokusering med hode-relaterte impulsresponser.

I **Kapittel 6** trekkes det konklusjoner rundt denne masteroppgaven, og det diskuteres hva som bør trekkes til videre forskning. Avslutningsvis foreslås det applikasjoner dette arbeidet kan resultere i.

Tillegg A inneholder et skriv om en metode brukt for å måle impulsresponser.

Tillegg B inneholder artikkelbidraget til NORSIG 2006.

Tillegg C inneholder programbeskrivelser av de viktigste programmene som ble laget til denne masteroppgaven. Programmene er funksjoner og skript som benyttes i Matlab™.

1.1 Relatert arbeid

Å kunne fokusere lyd i hørbar akustikk er forbundet med det å lage virtuelle lydlandskaper, og er avhengig av egenskapen til systemet å kunne gjenskape et ønsket lydfelt rundt en lytter[36]. Lydfokusering i rom er komplekst, grunnet rommets komplekse propageringsegenskaper. Disse propageringsegenskapene er knyttet til ekkoer og gjenklang. I dette delkapitlet ser vi raskt på arbeider som har en viss tilknytning til denne masteroppgaven.

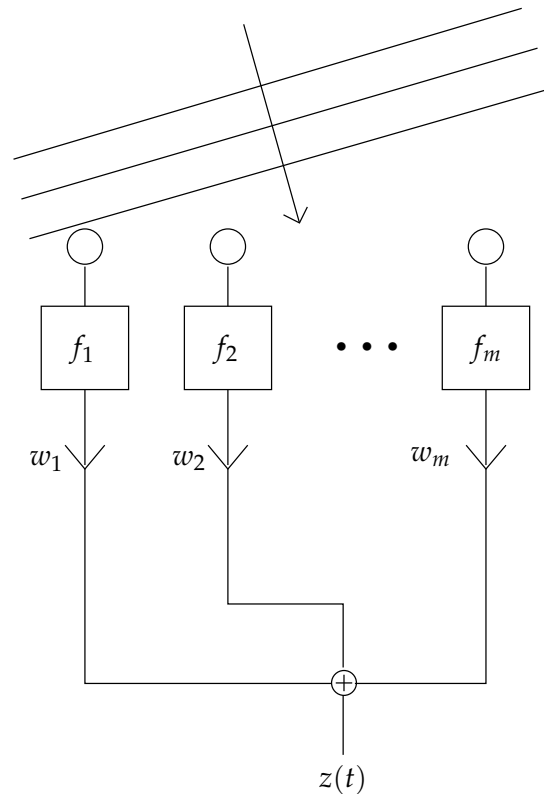
Denne oppgaven berører både kildeparasjon og lydfokusering, med hovedfokus på lydfokusering. Prinsippene vi ser på her er ikke direkte knyttet til arbeidet i denne masteroppgaven, men de bærer deler av den. Spesielt er det teoretiske grunnlaget for tidsreversering og faktiske løsninger på blinde separasjonsproblemer langt utenfor det teoretiske grunnlaget for dette arbeidet. Resultater fra fokusering med tidsreversering har vært mer som en inspirasjon for mitt arbeide.

1.1.1 Forsinkelse-og-sum-stråleformer

Forsinkelse- og sum-stråleforming er den enkleste måten å fokusere lyd på. For denne fokuseringsmetoden må man ha en oppstilling av to eller flere sensorer, satt sammen til en antenne. For hver inngangskanal påfører man en forsinkelse. Forsinkelsene velges i henhold til hvilken retning man vil fokusere lyden inn på; Fokuseringsretningen bestemmes av forsinkelsene. Om forsinkelsene velges riktig, vil summen av sensorenes opptak være konstruktiv med tanke på det ønskede signalet, det vil si, gi konstruktiv interferens. Samtidig vil den destruktivt summere de uønskede signalene, fra andre retninger. Dette resulterer i at det ønskede signalet forsterkes, samtidig som støy og uønskede signaler svekkes. Forsinkelse-og-sum stråleformer er definert som[19],

$$z(t) = \sum_{m=1}^M w_m y_m(t - f_m).$$

Hvor $m = 1 \dots M$ er antall sensorer og f_m er tidsforsinkelsen som er påføres inngangskanalene. Vektene w_m kan optimeres for å forbedre strålens form og sidelobe nivåer[19], for å gjøre hovedloben skarpere og sidelobene lavere. Sensorenes plassering om antennens origo kan også variere. Om sensorene plasseres langs x-aksen til antennen er den et lineært array. I figur 1.1 ser vi et slikt lineært array. Hvis forsinkelsene velges korrekt kan vinkelen til planbølgen detekteres, og dermed kan det beregnes hvilken retning bølgen kommer fra. Mer kompliserte konstellasjoner er mulige, som to- og



Figur 1.1: Forsinkelse-og-sum stråleformer

tredimensjonale antenner.

Forsinkelse-og-sum stråleformer kan også brukes til å fokusere lyd fra høyttalere til et punkt. Man må ha en oppstilling av to eller flere høyttalere, hvor utgangen til hver av høyttalerene forsinkes med en passende forsinkelse. Med et valgt fokuspunkt vil det være forskjellig avstand fra hver høyttaler til dette punktet. Forsinkelsene beregnes ut fra disse avstandsforskjellene for deretter å påføres de tilhørende utgangene. Signalene fra hver høyttaler vil dermed nå fokuspunktet ved samme tidspunkt[36].

1.1.2 Blind Kildeseparasjon (BSS)

Fokuseringsproblemet kan i en viss forstand relateres til blind kildeseparasjon. I et rom hvor mange mennesker snakker samtidig vil et menneske være i stand til å separere en kilde fra de andre kildene som er tilstede. Med BSS forsøker man å imitere denne menneskelige egenskapen ved hjelp av digital signalbehandling. Problemet er at man hverken kjenner kildene el-

ler propageringsegenskapene til miljøet som kildene propagerer gjennom, derav uttrykket *blind* kilde-separasjon. BSS problemet kan formuleres på følgende måte: Med L lydkilder i rommet, $s_1(t), s_2(t) \dots s_L(t)$, tas det opp-tak av dette lydfeltet med M mikrofoner. Dette gir opphav til målingene $x_1(n), x_2(n), \dots, x_M(n)$, hvor enhver $x_i(n)$ er

$$x_i(n) = \sum_{j=1}^L a_{i,j}(n) * s_j(n) \quad i = 1, 2, \dots, M \quad (1.1)$$

$a_{i,j}(n)$ er impulsresponsen signal j filtreres av i det det propagerer til mikrofon i . Oppgaven er å finne filtre $w_{i,j}$ som skiller de M kildene fra hverandre:

$$\hat{s}_j(n) = \sum_{i=1}^M x_i(n) * w_{i,j}(n) \quad j = 1, 2, \dots, L \quad (1.2)$$

Man må da estimere M filtre til å skille ut ett av de L signalene. For å skille ut alle L signalene trengs $M \times L$ filtre. Når miksingene har skjedd i ekkofrie rom kan modellen forenkles til

$$x_i(n) = \sum_{j=1}^L a_{ij} * s_j(n - \delta_{ij}) \quad i = 1, 2, \dots, M. \quad (1.3)$$

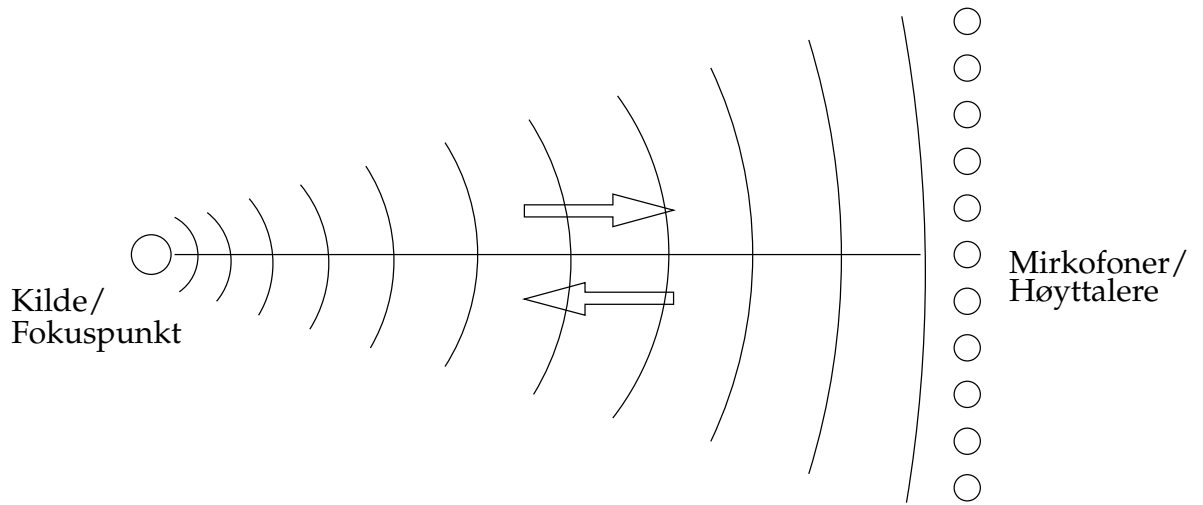
a_{ij} og δ_{ij} er henholdsvis dempningen og forsinkelsen fra kilde j til mikrofon i . Separasjonsfilterne som kun skal utligne $a_{i,j}(n)$ må estimeres fra $x_i(n)$ som er konvolusjonene av både $s_i(t)$ og $a_{i,j}(t)$, samplet.

En rekke metoder er utviklet for å løse slike problemer, noen foreslåtte metoder finnes i [3], [5] og [25]. Jeg skal ikke løse blinde kilde-separasjonsproblemer i denne oppgaven. Som vi skal se i kapittel 3 er det grunnleggende BSS-problemet formulert på omtrent samme måte som metodikken benyttes i denne oppgaven, som vil bli gjennomgått i neste kapittel.

1.1.3 Lydfokusering med tidsreversering

Lydfokusering med *tidsreversering* er forsket mye på innen ultralyd og undervannsakustikk. Denne fokuseringsmetoden baserer seg på tidsreverserings-invariansen til bølgelikningen. Den lineære bølgelikningen som modellerer propageringen av akustiske signaler gjennom et medium er en annen ordens differensiallikning i tid. Bølgeutbredelsen følger funksjonen[36]:

$$\nabla^2 p - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = 0 \quad (1.4)$$



Figur 1.2: Oppsett ved tidsreversering

Med tidsinvariansen menes at hvis $p(\mathbf{r}, t)$ er en løsning av 1.4 så er også $p(\mathbf{r}, -t)$ en løsning. Da vil det i teorien finnes et sett bølger som vil gå tilbake, langs de samme stiene som den opprinnelige bølgen fulgte, tilbake til punktet som bølgen ble sendt ut fra.

Tidsreversering er en prosess og kan beskrives ved følgende oppsett, en punktkilde s plasseres i et medium, et tidsreverseringsspeil plasseres med en avstand L fra denne punktkilden. Tidsreverseringsspeilet kan være en lineær antenne hvor elementene er i stand til både å motta og sende signaler. En bølge sendes fra punktkilden, og måles av elementene i tidsreverseringsspeilet. Målingene reverseres i tid, for deretter å sendes ut av tidsreverseringsspeilet. Når de tidsreverserte signalene sendes ut vil signalet omtrentlig propagere bakover til kilden, som om feltet gikk bakover i tid[36]. Til slutt måles feltet langs aksene der punktkilden er plassert, for å evaluere hvor godt lydfeltet er fokusert [33]

$$d(x) = \max_t(p(x, t)). \quad (1.5)$$

Der x er aksene det måles langs, $p(x, t)$ er lydstryken på det tidsreverserte feltet.

Tidsreversering i ikke-ekkoefrie rom

I [36] undersøkes lydfokusering basert på denne framgangsmåten, i et rom som ikke er lyddempet. Rommet inneholder møbler og datamaskiner, som vil avgi ytterligere ekkoer og refleksjoner. Når lyden sendes ut fra punktkilden vil denne lyden treffe veggene, gulv, tak og møblene i rommet.

Dette vil resultere i ekkoer av den opprinnelige lyden. Disse ekkoene vil reflekteres inn i den mottakende antennen. Når de tidsreverserte signalene sendes ut igjen av tidsreverseringsspeilet kan man tenke seg at ekkoene som kom fra veggene, tak, gulv og møbler sendes ut fra hver sin virtuelle antenne. Dette resulterer i at den effektive antennen i fokuspunktet øker[36]. Sammenliknet med fokusering i et fritt felt, vises det at rommets refleksjoner bidrar til økte fokuseringsegenskaper. I [36] vises det også at lydfokusering med tidsreversering gir bedre resultater enn med forsinkelse-og-sum fokusering.

Super-oppløsning

Å fokusere en bølge med tidsreversering i et homogent medium, vil gi et romlig fokuspunkt om punktkilden som er lik $\frac{\lambda L}{a}$. λ er bølgelengden til bæresignalet til pulsen som skal fokuseres, L er lengden fra punktkilden til tidsreverseringsspeilet og a er lengden til tidsreverseringsspeilet. I uordnede medier kan det romlig fokuspunktet være lik $\frac{\lambda L}{a_e}$. a_e er lengden på det effektive tidsreverseringsspeilet (som kan økes ved fokusering i uordnede medier), dette kalles *super-oppløsning*[4].

Det er gjennomført eksperimenter der super-oppløsning er observert i bølgeledere og i lukkede hulrom med veldig gode refleksjonsegenskaper. En bølgeleder kan for eksempel være to metallplater som lyden sendes gjennom, lyden reflekteres i disse sidene, og ledes gjennom oppsettet. I [8] vises det at super-oppløsningen observeres dersom det lukkede hulrommet har en kaotisk utforming. Informasjon man vanligvis ville ha mistet reflekteres inn imot sensorene igjen. Ettersom disse ekkoene kommer inn på et senere tidspunkt enn direktestrålen, må feltet måles over en lengre tidsperiode[8].

Tidsreversering i applikasjoner for hørbar lyd

Imidlertid nevnes det i [36] at denne metoden å fokusere lyd på ikke er særlig egnet for audio applikasjoner grunnet *temporære sideløber*. Disse har en effekt som gjør det problematisk å gjengi perkusive lyder og konsonanter som p og d særlig godt.

1.1.4 Lydfokusering med inversfiltrering

Å kunne kontrollere et lydfelt i et rom har vært av interesse for forskere fra mange disipliner, og kan formuleres som et lydfokuseringsproblem. Et alternativ til å bruke tidsreversering kan være å løse et inverst problem[37].

I denne masteroppgaven skal jeg se nærmere på det å kunne fokusere lyden til et bestemt punkt, og samtidig undertrykke andre punkter. Løsningen baserer seg på en metode som forklares i artikkelen "*Inverse Filtering og Room Acoustics*"[26]. Denne metoden er motivert gjennom å kunne finne eksakte inversfiltere til impulsresponsene til et rom.

Kapittel 2

Teori

I dette kapitlet vil jeg ta for meg grunnleggende teori for denne masteroppgaven. Innledningsvis ser jeg på minste kvadraters metode og singularverdidekomposisjonen til en matrise. Deretter ser jeg på noen egenskaper ved signaler i rom, lineære systemer, og måter å representere disse systemene på. Det å lage inverse filtere blir også gjennomgått. Til slutt utvides problemstillingen til multikanals inversfiltrering, i kontekst av lydfokusering og kildeseparasjon.

2.1 Minste kvadraters metode

Mange praktiske problemer innen blant annet statistikk og signalbehandling krever løsning av et likningsystem. Et lineært likningsystem i sin generelle form er et oppsett av m likninger og n ukjente variable

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + \dots + a_{1,n}x_n &= b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \dots + a_{2,n}x_n &= b_2 \\ &\vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \dots + a_{m,n}x_n &= b_m. \end{aligned} \tag{2.1}$$

Dette kan skrives om på matriseform

$$\mathbf{Ax} = \mathbf{b} \tag{2.2}$$

Hvor $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]^T$, $\mathbf{b} = [b_1 \ b_2 \ \dots \ b_m]^T$ og

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{bmatrix}$$

Jeg skal se nærmere på likningsystemer der antall likninger er større enn antall variable. Dette kalles *overbestemte* likningsystemer. Slike likningsystemer kan løses ved å bruke *minste kvadraters metode*

Anta at vi har en matriselikning som 2.2, hvor $\mathbf{A} \in \mathbb{R}^{m,n}$, med $m > n$, $\mathbf{b} \in \mathbb{R}^m$, og $\mathbf{x} \in \mathbb{R}^n$. Problemet består i å bestemme den ukjente vektoren \mathbf{x} . Om ikke \mathbf{b} er i kolonnerommet til \mathbf{A} , noe som ikke er tilfellet generelt, må løsningen estimeres[22]. Vi må da finne en “beste approksimasjon” fra \mathbf{Ax} til \mathbf{b} . For å finne denne, er det lettest å minimere kvadratet av den euklidske avstanden mellom \mathbf{Ax} og \mathbf{b} . Dette er det samme som å minimere uttrykket

$$\varepsilon = \|\mathbf{b} - \mathbf{Ax}\|_2^2 = \sum_{i=1}^m (b_i - \sum_{j=1}^n a_{ij}x_j)^2. \quad (2.3)$$

Minimum kan finnes ved å derivere funksjonen ε med hensyn på x_j og sette disse lik null. Dette resulterer i likningene

$$\begin{aligned} \frac{\partial \varepsilon}{\partial x_1} &= -2 \sum_{i=1}^m a_{i1} (b_i - \sum_{j=1}^n a_{ij}x_j) = 0 \\ &\vdots \\ \frac{\partial \varepsilon}{\partial x_n} &= -2 \sum_{i=1}^m a_{in} (b_i - \sum_{j=1}^n a_{ij}x_j) = 0 \end{aligned}$$

Satt på matriseform får vi

$$-2\mathbf{A}^T(\mathbf{b} - \mathbf{Ax}) = \mathbf{0} \implies \mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b} \quad (2.4)$$

Dette er de kjente *normallikningene*. Om $\mathbf{A}^T\mathbf{A}$ er ikke-singulær, er løsningen unik og gitt ved

$$\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} \quad (2.5)$$

2.1.1 Kolonnerom og rang

Kolonnerommet til en matrise $\mathbf{A} \in \mathbb{R}^{m,n}$ er mengden av alle lineære kombinasjoner av kolonnene til \mathbf{A} . Rangen til en matrise, $\text{rank}(\mathbf{A})$, er dimensjonen til kolonnerommet til \mathbf{A} [22]. Om en matrise $\mathbf{A} \in \mathbb{R}^{m,n}$ har rang mindre enn n , er matrisen av mangelfull rang. Dette betyr at ikke alle kolonnene til \mathbf{A} er lineært uavhengige, og \mathbf{A} inneholder derfor redundant informasjon. Om en matrise \mathbf{A} har full rang kalles matrisen $\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ den pseudo-inverse matrisen til \mathbf{A} [11]. Dersom en matrise \mathbf{A} har mangelfull rang kan den pseudo-inverse uttrykkes ved hjelp av singulærverdidekomposisjonen til \mathbf{A} .

2.2 Singulærverdidekomposisjon

Singulærverdidekomposisjonen av en matrise $\mathbf{A} \in \mathbb{R}^{m,n}$ er definert som[11]:

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{i=1}^n \mathbf{u}_i \sigma_i \mathbf{v}_i^T \quad (2.6)$$

Hvor $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_m) \in \mathbb{R}^{m,m}$ og $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_n) \in \mathbb{R}^{n,n}$ er matriser med orthonormale kolonner, $\mathbf{U}^T \mathbf{U} = \mathbf{V}^T \mathbf{V} = \mathbf{I}_n$. Den diagonale matrisen $\mathbf{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_p)$, $p = \min(m, n)$, inneholder de *singulære verdiene* til matrisen \mathbf{A} . Singulærverdiene er ordnet i ikke-økende rekkefølge,

$$\sigma_1 \geq \sigma_2 \geq \dots \sigma_p \geq 0.$$

Vektorene \mathbf{u}_i og \mathbf{v}_i kalles henholdsvis de venstre og høyre singulærvektorene til \mathbf{A} . Når singulærverdidekomposisjon utføres i kontekst av minstekvadraters problem, $m > n$, kan matrisene velges til å ha dimensjoner $\mathbf{U} \in \mathbb{R}^{m,n}$, $\mathbf{\Sigma} \in \mathbb{R}^{n,n}$ og $\mathbf{V} \in \mathbb{R}^{n,n}$. Denne singulærverdidekomposisjonen kalles "tynn" siden \mathbf{U} er rektangulær[13]. Om en matrise \mathbf{A} har mangelfull rang, $p < n$, er singulærverdiene gitt ved[11]

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq \sigma_{p+1} = \dots = \sigma_n = 0$$

Singulærverdiene kan gi indikasjoner om hvorvidt en matrise er nær ved å ha mangelfull rang. Dette er tilfellet om noen singulærverdier er veldig små. Dette indikerer at det finnes kolonner som numerisk sett ikke er lineært uavhengige.

En løsning av et dårlig kondisjonert system er veldig sensitiv med hensyn til pertubasjoner i både \mathbf{A} og \mathbf{b} . Løsningen \mathbf{x} kan uttrykkes ved hjelp av singulærverdidekomposisjonen til \mathbf{A} . På denne formen kan man lettere se problemene som oppstår.

2.3 SVD og minste kvadraters metode

En minste kvadraters løsning \mathbf{x} , av en matriselikning $\mathbf{A}\mathbf{x} = \mathbf{b}$, kan uttrykkes ved hjelp av singulærverdidekomposisjonen til \mathbf{A} . Jeg ser først på $\mathbf{A}^T \mathbf{A}$ ved singulærverdidekomposisjonen til \mathbf{A} ,

$$\mathbf{A}^T \mathbf{A} = \mathbf{V}\mathbf{\Sigma}\mathbf{U}^T \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \mathbf{V}\mathbf{\Sigma}^2 \mathbf{V}^T.$$

Den inverse til denne matrisen er gitt ved,

$$(\mathbf{A}^T \mathbf{A})^{-1} = (\mathbf{V}\mathbf{\Sigma}^2 \mathbf{V}^T)^{-1} = (\mathbf{V}^{-1})^T \mathbf{\Sigma}^{-2} \mathbf{V}^{-1} = \mathbf{V}\mathbf{\Sigma}^{-2} \mathbf{V}^T, \quad (2.7)$$

siden $\mathbf{V}^{-1} = \mathbf{V}^T$, $\mathbf{I}^{-1} = \mathbf{I}$ og $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$. Løsningen av normallikningene uttrykt gjennom singularverdidekomposisjonen er,

$$\begin{aligned}\mathbf{x} &= (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} = \mathbf{V} \mathbf{\Sigma}^{-2} \mathbf{V}^T \mathbf{V} \mathbf{\Sigma} \mathbf{U}^T \mathbf{b} \\ &= \mathbf{V} \mathbf{\Sigma}^{-1} \mathbf{U}^T \mathbf{b} = \sum_{i=1}^n \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \mathbf{v}_i.\end{aligned}\tag{2.8}$$

Vi ser at løsningen ikke vil gi mening om matrisen \mathbf{A} har mangelfull rang, det vil si en av singularverdiene er null. Om noen av singularverdiene er veldig små, kan visse komponenter/delbidrag til løsningen \mathbf{x} “blåses” opp av disse komponentene. Da kan normen til løsningen bli veldig stor, noe som kan by på problemer ved for eksempel en minste kvadraters tilpasning.

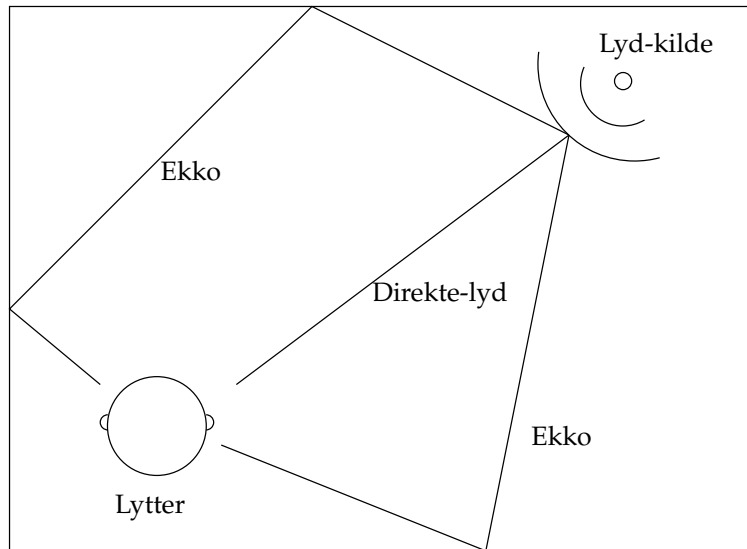
2.4 Litt om hvordan lyd oppfører seg i rom

Anta at vi har en lydkilde i et rom. Dersom det gjøres en måling av lyden i et punkt i rommet vil lyden gjennomgå en transformasjon fra kilden til målepunktet. Denne transformasjonen vil, i tillegg til avspillings- og opptaksutstyret, avhenge av størrelsen på rommet og hva slags vegger, tak og gulv rommet består av. På grunn av ekko og gjenklang vil lyden transformeres forskjellig i forhold til posisjonene til sender og mottaker. Hvis vi antar at gjengivelsesutstyret er perfekt, består denne transformasjonen av tre viktige deler: direkte-lyd, tidlige ekkoer og gjenklang. Disse tre egenskapene utgjør tilsammen det som kalles rommets *impulsrespons*[24].

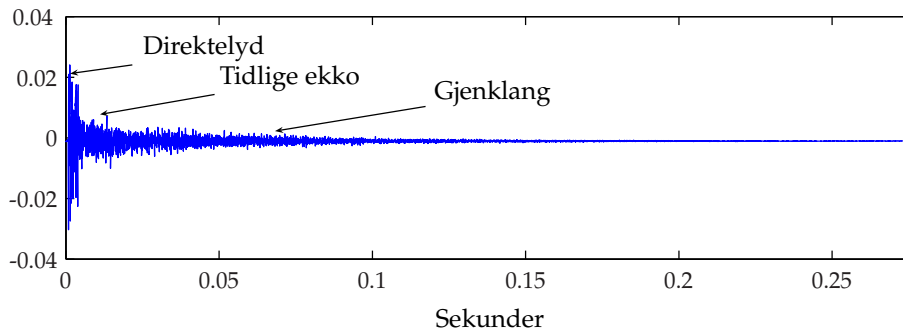
Direktelyden har den korteste banen fra kilden til lytteren og når lytteren først. De tidlige ekkoene når lytteren etter direktelyden. Deretter kommer de sene refleksjonene, de er tett pakket sammen, og har eksponensielt avtagende amplitude. Gjenklangstiden til et rom er definert som den tiden det tar til impulsresponsen har nådd et nivå under 60dB. Bildet i figur 2.1 viser et forenklet bilde av situasjonen beskrevet her. I virkeligheten kan det være tusenvis av refleksjoner som når lytteren. Det er imidlertid mange ting man kan ta med i betraktningen når man snakker om lydbølgeutbredelse i akustiske omgivelser. Noen helhetlig beskrivelse skal ikke utføres her. De viktigste egenskapene i denne oppgaven, er de som utgjør impulsresponsen. En impulsrespons målt i et rom er vist i figur 2.2. Pilene i figuren peker på de egenskapene nevnt i denne seksjonen.

2.5 Systemer og systemidentifikasjon

Det at signaler altereres eller gjennomgår en *transformasjon*, har man innen signalbehandling en generell betegnelse for: *Signalet har passert igjennom et*



Figur 2.1: Enkel modell av ekkoer i rom



Figur 2.2: En impulsrespons målt i et rom

system. Dette gjelder for signaler generelt enten det er snakk om digitale eller kontinuerlige akustiske signaler. Dersom vi begrenser oss til det digitale domenet vil et generelt lineært og tidsinvariant (LTI) system være på formen

$$y(n) = - \sum_{k=1}^N a_k y(n-k) + \sum_{k=0}^M b_k x(n-k). \quad (2.9)$$

Denne systemrepresentasjonen kalles en konstant-koeffisient differenslikning [32]. Systemet er et inngang-utgangs forhold som relater utgangen til inngangen gjennom en lineær transformasjon. Systemer kan også representeres gjennom dets *systemfunksjon* $H(z)$, og dens tidsdomene-ekvivalente *impulsrespons* $h(t)$. Systemfunksjonen beskriver systemets karakteristikk i

z-omenet, mens impulsresponsen beskriver karakteristikene i tidsdomenet. Forholdet mellom systemfunksjonen og impulsresponsen er gitt ved z-transformen[15]

$$H(z) = \sum_{n=-\infty}^{\infty} h(n)z^{-n} \quad (2.10)$$

For å relatere impulsresponsen til det generelle LTI-systemet i 2.9 kan man gå fram på følgende måte[32]. Vi tar z-transformen til 2.9

$$\begin{aligned} Y(z) &= - \sum_{k=1}^N a_k Y(z) z^{-k} + \sum_{k=0}^M b_k X(z) z^{-k} \\ Y(z) \left(1 + \sum_{k=1}^N a_k z^{-k}\right) &= X(z) \sum_{k=0}^M b_k z^{-k} \\ \frac{Y(z)}{X(z)} &= \frac{\sum_{k=0}^M b_k z^{-k}}{\left(1 + \sum_{k=1}^N a_k z^{-k}\right)} = H(z) \end{aligned}$$

Impulsresponsen kan man få ved å ta den inverse z-transformen til systemfunksjonen $H(z)$. Gitt impulsresponsen til et system, da er inngangs-utgangs forholdet gitt ved konvolusjons operatoren,

$$y(n) = h(n) * x(n) = \sum_{k=-\infty}^{\infty} h(k)x(n-k). \quad (2.11)$$

Jeg skal i denne oppgaven se på systemer der $a_k = 0$ i 2.9. Denne typen filtre er kjent som endelig-impulsrespons filtre (FIR-filtre).

I mange tilfeller er systemet ukjent og det er interessant å finne karakteristikene til dette systemet. Systemidentifikasjonsprosessen består i å bestemme disse, basert på et sett målinger utført på systemet. Gitt en inngangssekvens $x(n)$, stimulerer man systemet med denne inngangssekvensen, og anskaffer seg $y(n)$. Ved hjelp av $x(n)$ og $y(n)$, kan man bestemme impulsresponsen. Det er kjent at konvolusjon i tidsdomenet er det samme som multiplikasjon i z-omenet[32]. Hvis vi tar utgangspunkt i dette kan systemidentifikasjonsprosessen beskrives som:

$$Y(z) = H(z)X(z) \Rightarrow H(z) = \frac{Y(z)}{X(z)}$$

Systemidentifikasjon er en egen disiplin og et eget fagfelt. I denne oppgaven skal jeg ikke gå nærmere inn på generell systemidentifikasjon. Det som er av interesse for denne oppgaven er å kunne bestemme impulsresponsen til et system.

2.6 Menneskets hørsel som et system

Lyd gjennomgår en transformasjon når den propagerer fra et punkt til et annet i et rom. Når mennesker oppfatter lyd filtreres lyden ytterligere av overkropp, hode og ikke minst det ytre øret[12]. Denne transformasjonen representeres ved egne systemer, *hode-relaterte systemfunksjoner* eller *hode-relaterte impulsresponser*. Som for rom-impulsresponser vil disse impulsresponsene også avhenge av posisjonene til kilde og lytter.

2.6.1 Tidsforsinkelse og amplitudeforskjell til en innkommende bølge

Dersom et signal propagerer mot en lytter, og signalkilden ikke er tilstede rett foran, rett ovenfor eller rett bak lytteren vil det være forskjeller i signalene som mottas i de to ørene. Disse forskjellene benytter mennesker seg av for å lokalisere lydkilder[14].

Amplitudeforskjell

Amplitudeforskjell er nivåforskjeller som vil være tilstede i det mottatte signalet i hvert av ørene. Hvis en lydkilde er tilstede til høyre for lytteren vil hodet skygge for lyden som når det venstre øret, signalet mottatt i det venstre øret vil da bli noe dempet.

Tidsforsinkelse

Dersom et signal propagerer fra et punkt til høyre for lytteren, vil signalet nå det høyre øret på et tidligere tidspunkt enn det venstre. Da vil det være tilstede en tidsforsinkelse mellom signalene mottatt i de to ørene.

Diffraksjon

Diffraksjon er et fysisk fenomen som gjør at lydbølger vil propagere rundt objekter. Dette forekommer effektivt om bølgelengden til lydbølgen er betydelig større enn objektet [34].

Begrensinger ved lokalisering av lydkilder med tidsforsinkelse og amplitudeforskjell

Diffraksjon medfører at amplitudeforskjellen ikke vil gjøre seg spesielt gjeldene for frekvenser lavere enn 500Hz[14]. På samme måte som amplitudeforskjellen er av begrenset nytte for lokalisering gjelder det samme for tidsforsinkelsen. Ved å benytte tidsforsinkelsen mellom signalene mottatt i ørene kan det oppstå tvetydighetsproblemer. Tvetydighetsproblemene

oppstår når bølgelengden er sammenlignbar med lengden mellom målepunktene. Det kan da oppstå forvirring om hvorvidt lydkilden er plassert til høyre eller til venstre for lytteren, fordi bølgetoppene vil være tilstede ved begge ørene samtidig[14]. Tilsammen utgjør disse begrensningene et gap i frekvensområdet, der hverken amplitudeforskjellen eller tidsforsinkelsen i de mottatte signalene i ørene er tilstrekkelige for å lokalisere lydkilder.

Om man benytter disse forskjellene i signaler til å syntetisere et lydlandskap gjennom hodetelefoner, vil ikke lydene oppfattes som om de kom utenfra, men heller oppfattes som om eksisterte inne i hodet[14]. Forskning har vist at denne "inne i hodet"-effekten ikke er tilstede om man syntetiserer med hode-relaterte systemfunksjoner.



Figur 2.3: KEMAR-dukken

2.6.2 Hode-relaterte systemfunksjoner

Hode-relaterte systemfunksjoner er funksjoner av 4 variable, av posisjon x, y, z og frekvens f . Disse avhenger generelt av posisjonen lydkilden har i forhold til lytteren.

Hode-relaterte systemfunksjoner er komplette beskrivelser av de ytre filterne som mennesket lytter igjennom, fra en gitt posisjon. For å få en helhetlig beskrivelse av hvordan et menneske lytter, må man ha et sett med impulsresponser.

Det finnes både matematiske modeller av hode-relaterte systemfunksjoner og estimerer hvor man har målt og estimert ved hjelp av impulsresponsmålinger. Målingene utføres ved at en veldig liten mikrofon plasseres inne i øret på testpersoner. Signaler spilles av med en høyttaler som tas opp av mikrofonen. Høyttaleren flyttes for å måle impulsresponsene fra flere posisjoner rundt testpersonen. Målinger utføres også på modeller av mennesker, som for eksempel KEMAR-dukken, se figur 2.3. Denne dukken har et relativt høyt detaljnivå. Enklere modeller har også blitt brukt. Her er både hode og overkropp kuler[2]. Flere hode-relaterte systemfunksjonsdatabaser er tilgjengelige på internett hvor målingene er utført på mennesker og KEMAR-dukken, dette dokumenteres i henholdsvis [12] og [1].

2.7 Inversfiltering/dekonvolusjon

Når et signal har passert gjennom et system, vil det ofte være interessant å kunne gjenvinne dette signalet. Vi antar at denne transformasjonen består av en filtrering med et FIR-filter, filtreringen beskrives med konvolusjonsoperatoren 2.11. For å gjenvinne signalet må man reversere transformasjonen signalet har gjennomgått. Denne reverseringen består i å finne et filter, som utligner effekten av systemet signalet har passert gjennom. Slike filtre kalles inverse filtre. Inversfiltrering er formulert analytisk på følgende måte,

$$h(n) * g(n) = \delta(n) \quad \xrightarrow{\mathcal{Z}} \quad H(z)G(z) = 1. \quad (2.12)$$

Gitt et system representert ved dets impulsrespons $h(n)$, vil vi finne et filter $g(n)$ som reverserer filtreringen med $h(n)$. Anta at et signal $x(n)$ har passert gjennom et system og dermed har blitt konvolvert med $h(n)$, $y(n) = x(n) * h(n)$. Anta videre at det eksisterer et filter $g(n)$ som er slik at

$$y(n) * g(n) = x(n) * h(n) * g(n) = x(n) * \delta(n) = x(n)$$

$$\downarrow \mathcal{Z}$$

$$Y(z)G(z) = X(z)H(z)G(z) = X(z)$$

Da er $g(n)$ det inverse filteret til $h(n)$. $x(n)$ har blitt gjenvunnet eller dekonvolvert ut av $y(n)$.

Anta en impulsrespons $h(n)$, av endelig lengde N . Anta videre at $h(n)$ har et inverst filter $g(n)$, av lengde N . Da vil konvolusjonen av disse to signalene være på formen

$$\delta(n) = \sum_{k=0}^{N-1} h(n-k)g(k). \quad (2.13)$$

Denne konvolusjonssummen av endelig lengde kan settes på matriseform,

$$\begin{bmatrix} h(0) & & & & \\ & h(0) & & & \\ & \vdots & & & \\ & & \ddots & & \\ h(N-1) & & & h(0) & \\ & h(N-1) & & & \\ & & \ddots & & \\ & & & \ddots & \\ 0 & & & & h(N-1) \end{bmatrix} \begin{bmatrix} g(0) \\ \vdots \\ g(N-1) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (2.14a)$$

$$\mathbf{H}\mathbf{g} = \mathbf{d}. \quad (2.14b)$$

Matrisen $\mathbf{H} \in \mathbb{R}^{2N-1, N}$ kalles en konvolusjonsmatrise[15]. Vektor \mathbf{g} er av lengde N og dens komponenter er filter-koeffisientene til $g(n)$. \mathbf{d} er av lengde $2N-1$ og inneholder de diskrete signalverdiene til $\delta(n)$.

Generelt er det inverse filteret \mathbf{g} ukjent, og må estimeres. Formuleringen i 2.14 over gir opphav til et lineært likningsystem. Dimensjonene til konvolusjonsmatrisen \mathbf{H} gjør at likningsystemet blir overbestemt. I seksjon 2.1 så vi at et slikt system kan løses ved normallikningene. Løsningen \mathbf{g} inneholder filterkoeffisientene til $g(n)$. $g(n)$ konvolvert med $h(n)$, vil gi en best mulig approksimasjon, $\hat{\delta}(n)$ i minste kvadraters forstand, til den ønskede responsen $\delta(n)$. Feilen mellom den ønskede responsen og den estimerte responsen, er gitt ved

$$\varepsilon = \|\mathbf{d} - \mathbf{H}\mathbf{g}\|_2^2$$

Valg av respons

Responsvektoren \mathbf{d} velges i henhold til definisjonen av inversfiltrering, men kan i prinsippet velges fritt [15]. Dette medfører at enhver ønsket respons kan estimeres ved at denne responsen settes som \mathbf{d} i matriselikningen. Ønskede responser kan være impulsreponser fra gode konserthaller, lavpassfiltere, høypassfiltere og så videre. Minste kvadraters metode også kan benyttes til å estimere impulsresponsen til et system.

Forsinkelse i tidsdomenet

Det har vist seg at resultatene av inversfiltrering kan forbedres om man innfører en heltallig forsinkelse l , i den ønskede responsen $\delta(n)$. Dette er vist i [29] ved hjelp av funksjoner i z -domenet. Den ønskede responsen i vektornotasjon, med forsinkelse, vil da være på formen

$$\mathbf{d} = [0 \cdots 0 \ 1 \ 0 \cdots 0]^T \quad (2.15)$$

Indekseringen til 1-tallet korresponderer med den heltallige forsinkelsen l . Resultatet av inversfiltreringen med tidsforsinkelse er

$$h(n) * g(n) = \delta(n - l) \quad (2.16)$$

Lengden til det inverse filteret

Lengden på det inverse filteret $g(n)$ kan velges å være både kortere og lengre enn $h(n)$. Resultatene kan ofte bedres dersom det inverse filteret velges lengre [15]. Om det inverse filteret velges til å ha lengde K vil konvolusjonsmatrisen \mathbf{H} ha dimensjonene $(N + K - 1, K)$, og vektoren \mathbf{d} , som inneholder den ønskede responsen vil være av lengde $N + K - 1$.

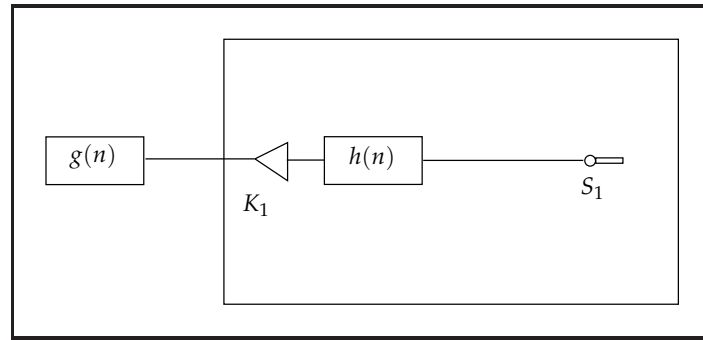
2.8 Anvendelse av inversfiltrering

Det er spesielt to anvendelser av inversfiltrering som er av interesse i denne oppgaven. Den ene anvendelsen er indikert i 2.7, å gjenvinne et signal som har passert gjennom et system. Den andre anvendelsen er å fokusere lyden mot et bestemt punkt i rommet. Da går man fram på følgende måte: Først estimerer man impulsresponsen i dette punktet. Deretter må det estimeres et invert filter til denne impulsresponsen. Til slutt pre-filtreres signalet før det sendes ut av høyttaleren. For å gjenvinne et signal er framgangsmåten noe annerledes. Anta at et signal sendes ut i et rom, og tas opp av en mikrofon. Anta videre at impulsresponsen til systemet signalet filtreres av er kjent. Det må så estimeres et invert filter til denne impulsresponsen. Til slutt filtreres opptaket fra mikrofonen med det inverse filteret. Fokuserings- og separasjonsproblemet er relatert til disse to anvendelsene av inversfiltrering.

2.9 Utvidet inversfiltrering

Inversfiltrering kan brukes til å fokusere lyden mot et punkt i rommet. I det tradisjonelle oppsettet brukes det en høyttaler og en mikrofon til dette formålet, se figur 2.4.

Generelt sett er det ikke mulig å realisere et eksakt inverst filter til denne impulsresponsen. I [26] foreslås en metode som overkommer denne begrensningen. Metoden kalles MINT (Multiple input/output Inverse Theorem). Som navnet på metoden indikerer, dreier det seg om å ta i bruk flere høyttalere og/eller mikrofoner for å løse et inverst problem. For å utligne romresponsen i ett punkt, brukes to høyttalere i stedet for én. For å gjenvinne et signal, brukes to mikrofoner. Man kan også bruke flere mikrofoner eller høyttalere.



Figur 2.4: Tradisjonelt oppsett ved invertering av impulsrespons

Anta et oppsett med to høyttalere K_1 og K_2 , og en mikrofon S_1 . Kall impulsresponsen fra K_1 til S_1 , $h_{1,1}(n)$, og impulsresponsen fra K_2 til S_1 $h_{1,2}(n)$, se figur 2.5. Anta at $h_{1,1}(n)$ og $h_{1,2}(n)$ er av endelig lengde N , og at det finnes to filtere $g_1(n)$ og $g_2(n)$ som er slik at

$$h_{1,1}(n) * g_1(n) + h_{1,2}(n) * g_2(n) = \delta(n), \quad (2.17)$$

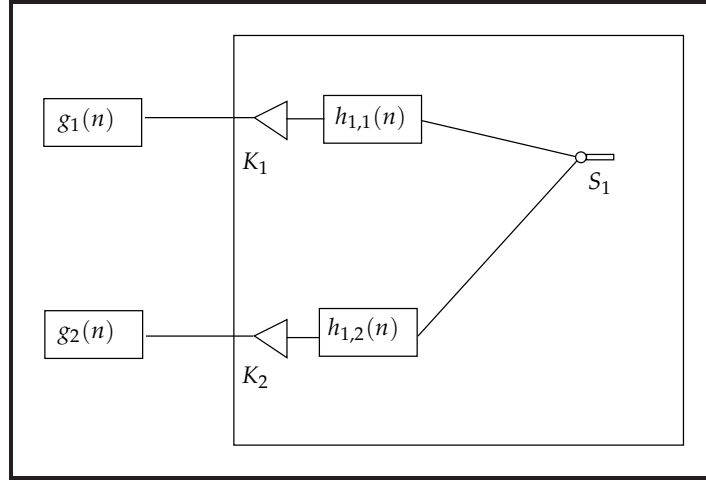
Gitt filterne $g_1(n)$ og $g_2(n)$, da vil vi ha utlignet romresponsen i punktet hvor $h_{1,1}(n)$ og $h_{1,2}(n)$ ble målt. Filterne $g_1(n)$ og $g_2(n)$ er ukjente, og må estimeres. Vi kan gjøre som i seksjon 2.7, og sette dette på matriseform [26],

$$\mathbf{H}_{1,1} \mathbf{g}_1 + \mathbf{H}_{1,2} \mathbf{g}_2 = \mathbf{d}. \quad (2.18)$$

Skriver vi om dette til blokkform får vi

$$[\mathbf{H}_{1,1} \ \mathbf{H}_{1,2}] \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix} = \mathbf{d}.$$

Mer spesifikt får vi et likningsystem på formen



Figur 2.5: Systemoppsett ved invertering av impulsrespons med to høyttalere og en mikrofon.

$$\begin{bmatrix}
 h_{1,1}(0) & & & h_{1,2}(0) & & \\
 \vdots & & & \vdots & & \\
 & & 0 & & & 0 \\
 \vdots & \ddots & & \vdots & \ddots & \\
 h_{1,1}(N-1) & & h_{1,1}(0) & h_{1,2}(N-1) & & h_{1,2}(0) \\
 & & \vdots & & & \vdots \\
 0 & \ddots & \vdots & 0 & \ddots & \vdots \\
 & & h_{1,1}(N-1) & & & h_{1,2}(N-1)
 \end{bmatrix}
 \begin{bmatrix}
 g_1(0) \\
 \vdots \\
 g_1(N-1) \\
 g_2(0) \\
 \vdots \\
 g_2(N-1)
 \end{bmatrix}
 =
 \begin{bmatrix}
 0 \\
 \vdots \\
 0 \\
 1 \\
 0 \\
 \vdots \\
 0
 \end{bmatrix}$$

Dette er et likningsystem på samme form som i forrige seksjon, og kan løses ved normallikningene. Løsningen er gitt ved,

$$\begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix} = \left(\begin{bmatrix} \mathbf{H}_{1,1}^T \\ \mathbf{H}_{1,2}^T \end{bmatrix} [\mathbf{H}_{1,1} \mathbf{H}_{1,2}] \right)^{-1} \begin{bmatrix} \mathbf{H}_{1,1}^T \\ \mathbf{H}_{1,2}^T \end{bmatrix} \mathbf{d}$$

$$\Rightarrow \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{1,1}^T \mathbf{H}_{1,1} & \mathbf{H}_{1,1}^T \mathbf{H}_{1,2} \\ \mathbf{H}_{1,2}^T \mathbf{H}_{1,1} & \mathbf{H}_{1,2}^T \mathbf{H}_{1,2} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{H}_{1,1}^T \\ \mathbf{H}_{1,2}^T \end{bmatrix} \mathbf{d}.$$

Anta at vi velger lengden på $g_1(n)$ og $g_2(n)$ til å være $N - 1$. Da vil den sammensatte konvolusjonsmatrisen $\mathbf{H} = [\mathbf{H}_{1,1} \mathbf{H}_{1,2}]$ få dimensjoner $(N + (N - 1) - 1, 2(N - 1)) = (2N - 2, 2N - 2)$. Da blir \mathbf{H} en kvadratisk matrise. Sett i en lineær algebra kontekst vil det være mulig å realisere en eksakt inverterbar matrise \mathbf{H}^{-1} , om \mathbf{H} har full rang.

Anta at et signal $x(n)$ har blitt filtrert av impulsresponser $h_{1,1}(n)$ og $h_{1,2}(n)$. Dette resulterer i to uavhengige realisasjoner av $x(n)$.

$$y_1(n) = x(n) * h_{1,1}(n)$$

$$y_2(n) = x(n) * h_{1,2}(n)$$

Disse to kan kombineres til å dekonvolvare $x(n)$ ut av $y_1(n)$ og $y_2(n)$. Gitt de inverse filterne $g_1(n)$ og $g_2(n)$, da kan $x(n)$ gjenvinnes ved,

$$\begin{aligned} & y_1(n) * g_1(n) + y_2(n) * g_2(n) \\ &= x(n) * h_{1,1}(n) * g_1(n) + x(n) * h_{1,2}(n) * g_2(n) \\ &= x(n)(h_{1,1}(n) * g_1(n) + h_{1,2}(n) * g_2(n)) \\ &= x(n) * \delta(n - l) \\ &= x(n - l). \end{aligned}$$

I følge simuleringer i [26] gir denne metoden generelt bedre resultater enn i det tradisjonelle oppsettet, utenom noen symmetriske posisjoner av mikrofon og høyttalere.

2.9.1 Undertrykkelsesfiltere

Inversfiltrering kan brukes til å invertare romresponsen i et punkt i rommet. For å relatere dette til fokuseringsproblemet må vi utvide problemet. Løsningsfilterne må også kunne undertrykke gitte punkter. Betrakt derfor et litt annet problem. Anta at vi har målt impulsresponsene fra to høyttalere K_1 og K_2 i et punkt vi vil undertrykke. La impulsresponsen fra K_1 til dette punktet være $h_{2,1}(n)$ og impulsresponsen fra K_2 til dette punktet være $h_{2,2}(n)$. Minimeringsproblemet som oppstår er,

$$\min_{\mathbf{g} \neq \mathbf{0}} \|[\mathbf{H}_{2,1} \ \mathbf{H}_{2,2}]\mathbf{g} - \mathbf{0}\|_2^2. \quad (2.19)$$

Hvor $\mathbf{0}$ er nullvektoren, en vektor hvor alle elementene er lik 0. Vi trenger beskrankningen $\mathbf{g} \neq \mathbf{0}$ for å unngå den trivielle null-løsningen, $\mathbf{g} = \mathbf{0}$. Løsningen vi er ute etter er filtere $g_1(n)$ og $g_2(n)$ som undertrykker et signal som filtreres av impulsresponsene $h_{2,1}(n)$ og $h_{2,2}(n)$. Anta at vi har løst minimeringsproblemet 2.19, løsningen \mathbf{g} inneholder to filtere $g_1(n)$ og $g_2(n)$. Om et signal filtreres med $g_1(n)$ og $g_2(n)$ før det sendes ut i rommet med de to høyttalerne, vil lyden være undertrykt i punktet hvor $h_{2,1}(n)$ og $h_{2,2}(n)$ ble målt. Undertrykkelsesproblemet skal i denne oppgaven ikke løses enkeltstående, men løses i kombinasjon med inversfiltreringsproblemet.

2.9.2 Kombinert inversfiltrering og undertrykkelse

For kombinert inversfiltrering og undertrykking formuleres følgende problem. Anta at vi har et punkt i rommet hvor vi vil invertere romresponsen. Anta videre at vi har ytterligere ett punkt som vi vil undertrykke. La oppsettet bestå av to høyttalere K_1 og K_2 , og to mikrofoner S_1 og S_2 . La impulsresponsene fra K_1 og K_2 til S_1 være henholdsvis $h_{1,1}(n)$ og $h_{1,2}(n)$, og impulsresponsene fra K_1 og K_2 til S_2 være henholdsvis $h_{2,1}(n)$ og $h_{2,2}(n)$. Dette gir opphav til følgende likningsystem,

$$\begin{bmatrix} \mathbf{H}_{1,1} & \mathbf{H}_{1,2} \\ \mathbf{H}_{2,1} & \mathbf{H}_{2,2} \end{bmatrix} \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{d} \\ \mathbf{0} \end{bmatrix} \quad (2.20a)$$

$$\mathbf{H}\mathbf{g} = \mathbf{d} \quad (2.20b)$$

hvor \mathbf{d} er som i 2.14. Med dette oppsettet vil det ikke lenger være mulig å få en kvadratisk matrise \mathbf{H} . La de inverse filterne $g_1(n)$ og $g_2(n)$ velges til å være av lengde K . Den sammensatte konvolusjonsmatrisen \mathbf{H} vil da få dimensjonene $(2(N + K - 1), 2K) = (2N + 2K - 2, 2K)$, som alltid gir opphav til et overbestemt likningsystem, med mindre $N = 1$.

Løsningen av denne matriselikningen \mathbf{g} inneholder koeffisientene til filterne $g_1(n)$ og $g_2(n)$. Dersom $g_1(n)$ og $g_2(n)$ benyttes til å prefiltrere henholdsvis utgangene K_1 og K_2 vil lyden være fokusert i S_1 , og samtidig være undertrykt i punktet S_2 , i minste kvadraters forstand.

$$S_1 = h_{1,1}(n) * g_1(n) + h_{1,2}(n) * g_2(n) \approx \delta(n - l)$$

$$S_2 = h_{2,1}(n) * g_1(n) + h_{2,2}(n) * g_2(n) \approx 0(n).$$

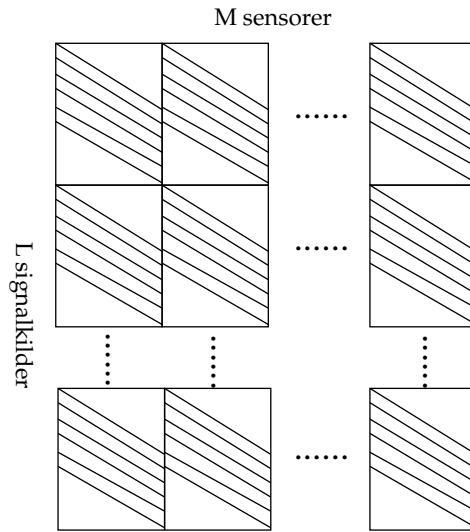
Hvor $0(n)$ er en sekvens med bare nuller av lengde $2N - 1$, om de inverse filterne $g_1(n)$ og $g_2(n)$ har lengde N .

Vi kan utvide problemet ytterligere ved å tilføye ennå et punkt vi vil undertrykke. Anta at dette punktet betegnes med S_3 , og la impulsresponsene fra høyttalerne K_1 og K_2 til S_3 være henholdsvis $h_{3,1}(n)$ og $h_{3,2}(n)$. Det utvidede likningsystemet blir da

$$\begin{bmatrix} \mathbf{H}_{1,1} & \mathbf{H}_{1,2} \\ \mathbf{H}_{2,1} & \mathbf{H}_{2,2} \\ \mathbf{H}_{3,1} & \mathbf{H}_{3,2} \end{bmatrix} \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{d} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \quad (2.21a)$$

$$\mathbf{H}\mathbf{g} = \mathbf{d}. \quad (2.21b)$$

Systemet kan utvides til å bruke vilkårlig mange sensorer/signalkilder, for kilde-separasjon[16], L signalkilder og M sensorer, L sensorer og M høyttalere for lydfokusering. Likningsystemet blir da



Figur 2.6: Matrisestruktur for kilde separasjon med M sensorer og L signalkilder.

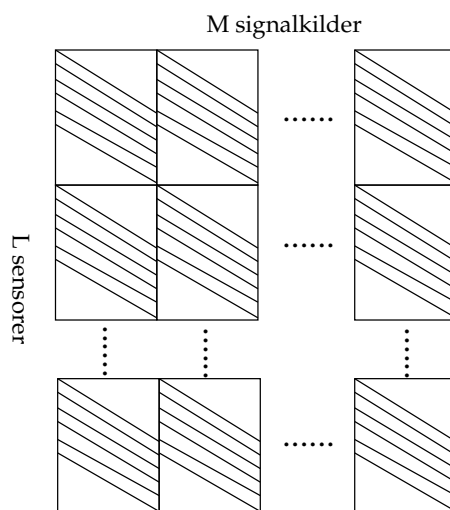
$$\begin{bmatrix} \mathbf{H}_{1,1} & \mathbf{H}_{1,2} & \cdots & \mathbf{H}_{1,M} \\ \mathbf{H}_{2,1} & \mathbf{H}_{2,2} & \cdots & \mathbf{H}_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{L,1} & \mathbf{H}_{L,2} & \cdots & \mathbf{H}_{L,M} \end{bmatrix} \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \vdots \\ \mathbf{g}_M \end{bmatrix} = \begin{bmatrix} \mathbf{d} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \quad (2.22a)$$

$$\mathbf{H}\mathbf{g} = \underline{\mathbf{d}} \quad (2.22b)$$

I figurene 2.6 og 2.7 vises båndstrukturen til matrisen \mathbf{H} for henholdsvis kilde separasjon og lydfokusering.

Eksakt løsbarehet

Anta nå at vi vil fokusere lyden mot et bestemt punkt, samt undertrykke andre punkter. Anta videre at vi har et system med L mikrofoner, og M høyttalere, og at vi måler impulsresponsene fra de M høyttalerne i de L punktene. Dette gir opphav til $M \times L$ impulsresponser. Da vil den sammensatte konvolusjonsmatrisen \mathbf{H} få dimensjoner $(L(N + K - 1), MK)$, hvor N er lengden på impulsresponsene $h_{i,j}(n)$ og K er lengden på de inverse filterene $g_j(n)$ for $i = 1, \dots, L$ og $j = 1, \dots, M$. Om ikke $\underline{\mathbf{d}}$ er i kolonnenrommet til \mathbf{H} er det kun i tilfellet der vi har flere høyttalere enn målepunkter vi kan løse fokuseringsproblemet eksakt. Dette vises i [16]. Jeg skal i denne oppgaven konsentrere meg om tilfellet der $L \geq M$, som vil gi opphav til overbestemte likningsystemer.



Figur 2.7: Matrisestruktur for lydfokusering med M høyttalere og L målepunkter/sensorer.

2.10 Kildeseparasjon og lydfokusering

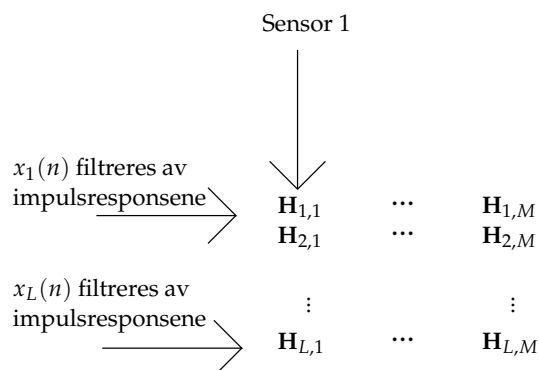
Metoden som er blitt presentert i de foregående seksjonene kan brukes til både kildeseparasjon og lydfokusering. Lydfokusering og kildeseparasjon deler et matematisk rammeverk som kan benyttes for å sette opp likningssystemer som modellerer problemstillingene. Målene/kriteriene som brukes for å evaluere om fokuseringen eller separasjonen er vellykket er imidlertid noe forskjellig, se delkapittel 2.11. Vi skal her se hvordan disse to situasjonene ser ut i sin generelle form.

2.10.1 Kildeseparasjon

For kildeseparasjon kan én sensors opptak gis ved den konvolutive summen (jamfør blind kildeseparasjon 1.1.2)

$$y_j(n) = \sum_{i=1}^L h_{i,j}(n) * x_i(n) \quad j = 1, 2, \dots, M. \quad (2.23)$$

Vi har L signaler $x_i(n)$, $i = 1, 2, \dots, L$, som vi vil skille ut fra de M konvolutive miksene. Det vil da finnes M optimale filtere $g_j(n)$, $j = 1, 2, \dots, M$, i minste kvadraters forstand, for hvert av de L signalene. For å separere ut de L signalene trenger vi $M \times L$ filtere[16]. Anta at vi skal skille signalet $x_1(n)$ fra de øvrige $L - 1$ signalene. La $x_1(n)$ være det signalet som filtreres av impulsresponsene $h_{1,1}(t), \dots, h_{1,M}(t)$. (jeg antar at disse impulsresponsene er kjent og at de er tilgjengelig i samplede versjoner). Konvolusjonsmatrisene $\mathbf{H}_{1,1}, \dots, \mathbf{H}_{1,M}$ tilhørende disse impulsresponsene settes


 Figur 2.8: Hjelpesfigur til oppsettet av matrisen \mathbf{H} ved kildeparasjon.

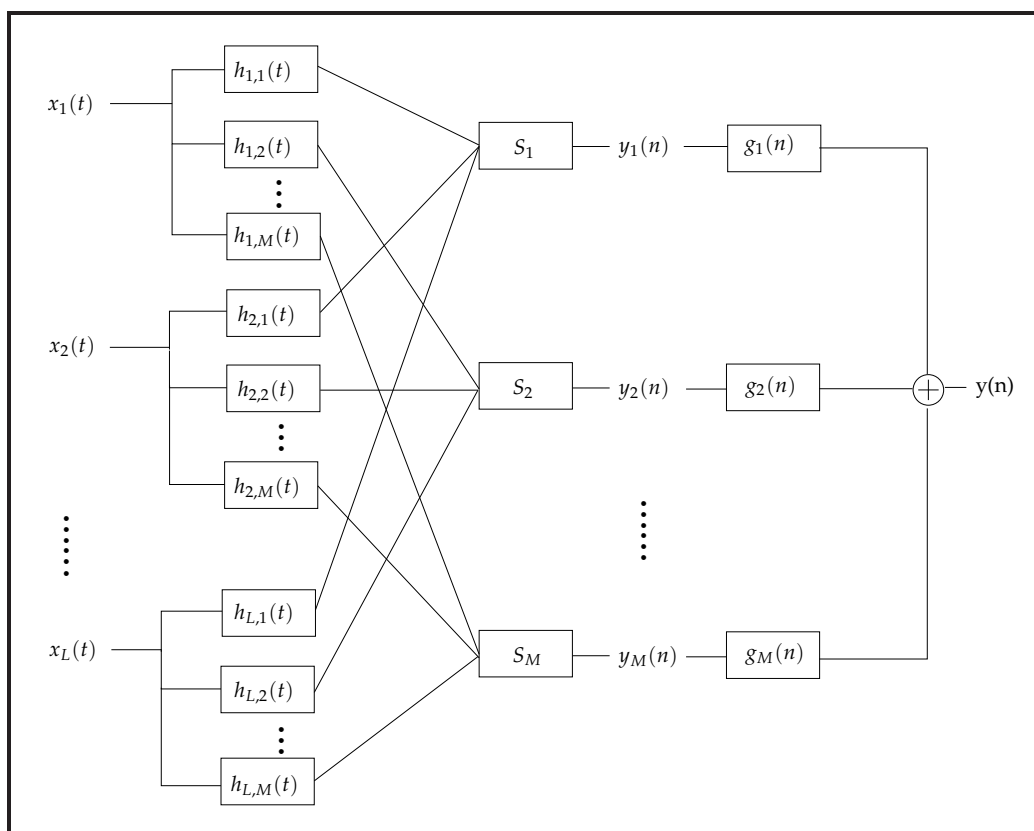
inn i øverste blokk-rad i den sammensatte konvolusjonsmatrisen \mathbf{H} , se figur 2.8. Konvolusjonsmatrisene tilhørende impulsresponsene kilden $x_2(n)$ filtreres av, $\mathbf{H}_{2,1}, \dots, \mathbf{H}_{2,M}$, settes inn i andre blokk-rad i \mathbf{H} . Tilsvarende gjøres for konvolusjonsmatrisene tilhørende impulsresponsene de øvrige kildene $x_3(n), \dots, x_L(n)$ filtreres av. \mathbf{H} settes så inn i likningsystemet 2.22 og likningsystemet løses ved normallikningene. Løsningen \mathbf{g} inneholder filterkoeffisientene til de M filterne $g_1(n), \dots, g_M(n)$ som brukes til å filtrere hver av sensorenes opptak,

$$y(n) = \sum_{j=1}^M y_j(n) * g_j(n). \quad (2.24)$$

Vi setter 2.23 inn i 2.24 og får:

$$\begin{aligned} y(n) &= \sum_{j=1}^M y_j(n) * g_j(n) \\ &= \sum_{j=1}^M \left(\sum_{i=1}^L x_i(n) * h_{i,j}(n) \right) * g_j(n) \\ &= \sum_{i=1}^L x_i(n) * \left(\sum_{j=1}^M h_{i,j}(n) * g_j(n) \right) \\ &= x_1(n) * d_1(n) + \sum_{i=2}^L x_i(n) * d_i(n) \\ &\approx x_1(n) * \delta(n-k) + \sum_{i=2}^L x_i(n) * 0(n) \end{aligned}$$

\mathbf{d}_i er i'te delvektor av den estimerte reponsen $\mathbf{H}\mathbf{g} = \hat{\mathbf{d}}$ og inneholder responsen $d_i(n)$ som er en approksimasjon til en sekvens med bare



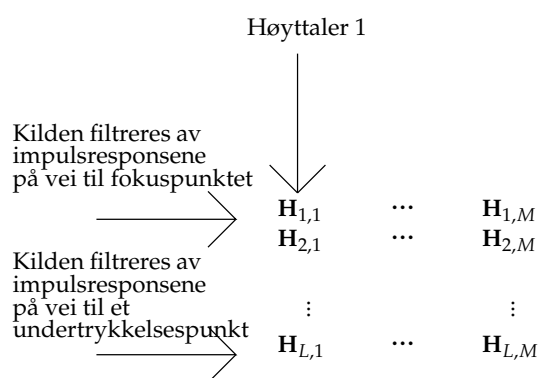
Figur 2.9: Systemrepresentasjon av kilde separasjon med L kilder og M mikrofoner, oppsettet viser separasjon av en kilde.

nuller. Ved utgangen $y(n)$ vil $x_1(n)$ være skilt ut fra de andre signalene $x_2(n), \dots, x_L(n)$. I figur 2.9 ser vi en systemrepresentasjon for dette. Dersom vi ville skille ut signalet $x_2(n)$ fra de andre signalene kunne man plassert konvolusjonsmatrisene tilhørende impulsresponsene $h_{2,1}(n), \dots, h_{2,M}(n)$ i øverste blokk-rad i \mathbf{H} og løst likningsystemet 2.22 på nytt. For å skille ut alle de L signalene fra hverandre ville vi ha måttet løst likningsystemet 2.22 L ganger ved å sette konvolusjonsmatrisene tilhørende impulsresponsene som disse signalene filtreres av i øverste blokk-rad i \mathbf{H} .

2.10.2 Lydfokusering

Dersom metoden anvendes til å fokusere lyd, er prinsippet likt, men framgangsmåten noe annerledes. I stedet for å separere en kilde fra en miks av flere, vil vi fokusere et signal $x(n)$ fra lydkilder til et punkt. La filterne $h_{1,1}(n), \dots, h_{1,M}(n)$ være samplede versjoner av impulsresponsene som beskriver filterne signalet $x(n)$ filtreres av, fra de M lydkildene, på vei til

fokuspunktet S_1 . Konvolusjonsmatrisene tilhørende disse impulsresponsene $\mathbf{H}_{1,1}, \dots, \mathbf{H}_{1,M}$ settes inn i øverste blokk-rad i den sammen satte konvolusjonsmatrisen \mathbf{H} , se figur 2.10. La impulsresponsene $h_{2,1}(n), \dots, h_{2,M}$ beskrive filterne kilden filtreres av på vei til et av punktene vi vil undertrykke, S_2 . Konvolusjonsmatrisene tilhørende disse impulsresponsene $\mathbf{H}_{2,1}, \dots, \mathbf{H}_{2,M}$ settes inn i andre blokk-rad i matrisen \mathbf{H} . Tilsvarende gjøres for konvolusjonsmatrisene tilhørende impulsresponsene som beskriver filterne signalet filtreres av på vei til de $L - 2$ andre punktene vi vil undertrykke, S_3, \dots, S_L .

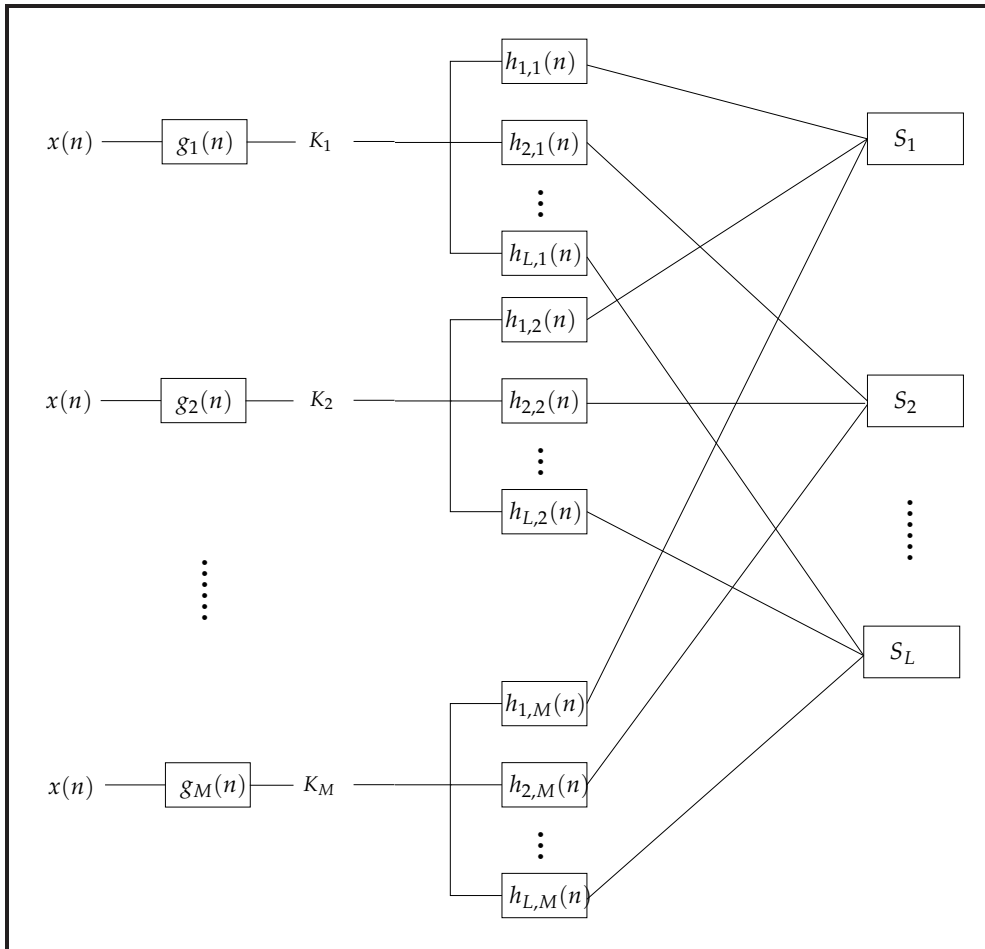


Figur 2.10: Hjelpfigur til oppsettet av matrisen \mathbf{H} ved lydfokusering.

\mathbf{H} settes så inn i likningssystemet 2.22, og likningsystemet løses. Løsningen \mathbf{g} inneholder filterkoeffisientene til de M filterne $g_1(n), \dots, g_M(n)$, som brukes til å prefiltrere de M utgangene.

Vi får:

$$\begin{aligned}
 S_1 &= \sum_{j=1}^M h_{1,j}(n) * g_j(n) * x(n) = \left(\sum_{j=1}^M h_{1,j}(n) * g_j(n) \right) * x(n) \approx \delta(n - k) * x(n) \\
 S_2 &= \sum_{j=1}^M h_{2,j}(n) * g_j(n) * x(n) = \left(\sum_{j=1}^M h_{2,j}(n) * g_j(n) \right) * x(n) \approx 0(n) * x(n) \\
 &\vdots \\
 S_L &= \sum_{j=1}^M h_{L,j}(n) * g_j(n) * x(n) = \left(\sum_{j=1}^M h_{L,j}(n) * g_j(n) \right) * x(n) \approx 0(n) * x(n)
 \end{aligned} \tag{2.25}$$



Figur 2.11: Systemrepresentasjon av lydfokusering med M høyttalere og L mikrofoner.

I figur 2.11 ser vi systemrepresentasjonen for dette. Om likningsystemet 2.22 lar seg løse eksakt vil vi kunne høre $x(n)$ i punktet S_1 og det vil være stille i de øvrige punktene S_2, \dots, S_L .

2.11 Separasjonskriterier

2.11.1 Kriterier for å måle om hvorvidt lydfokuseringen er vellykket

Anta at vi har ett oppsett med to høyttalere K_1 og K_2 og to mikrofoner S_1 og S_2 . La impulsresponsene fra K_1 og K_2 til S_1 være henholdsvis $h_{1,1}(n)$ og $h_{1,2}(n)$ og impulsresponsene fra K_1 og K_2 til S_2 være henholdsvis $h_{2,1}(n)$ og

$h_{2,2}(n)$. Anta at vi vil fokusere lyden fra K_1 og K_2 mot punktet S_1 , samtidig som vi undertrykker lyden i punktet S_2 . De optimale filterne $g_1(n)$ og $g_2(n)$ finnes ved å løse likningsystemet 2.20. Etter endt inversfiltrering vil disse to punktene gi opphav til hver sin respons

$$d_1(n) = h_{1,1}(n) * g_1(n) + h_{1,2}(n) * g_2(n) \approx \delta(n - l)$$

$$d_2(n) = h_{2,1}(n) * g_1(n) + h_{2,2}(n) * g_2(n) \approx 0(n)$$

I denne oppgaven brukes det to kriterier for å kunne måle hvor god fokusering som oppnås etter endt inversfiltrering. Ett av kriteriene er, den totale energien til responsen i fokuspunktet, relativt til den totale energien til responsen i ett undertrykkelsespunkt. Energien til et signal $x(n)$ av lengde N er gitt ved [32],

$$e = \sum_{n=1}^N |x(n)|^2 = \|x(n)\|_2^2. \quad (2.26)$$

Et relativt energinivå mellom responsen i fokuspunktet og undertrykkelsespunktet er gitt ved,

$$f_2(d_1(n), d_2(n)) = 10 \log_{10} \frac{\|d_1(n)\|_2^2}{\|d_2(n)\|_2^2}. \quad (2.27)$$

Her målt i desibel. Det måler energiforholdet mellom responsen i fokuspunktet, $d_1(n)$ og responsen i undertrykkelsespunktet $d_2(n)$. Dersom vi har et oppsett hvor det er flere undertrykkelsespunkter kan det også være interessant å se på energiforholdet mellom fokuspunktet og de øvrige undertrykkelsespunktene.

Dette kriteriet kan tolkes som den relative styrken lyden har i to posisjoner. Desto større denne verdien er, jo større forskjell i lydstyrke vil det være mellom fokuspunktet og undertrykkelsespunktet. Lyden vil kunne høres bedre i fokuspunktet enn i undertrykkelsespunktet. Dersom $f_1(d_1(n), d_2(n)) = 0$, vil det ikke være noen forskjell i lydstyrke mellom fokuspunktet og undertrykkelsespunktet. Jeg bruker også det relative maksimale utslaget i amplitude mellom fokuspunkt og et undertrykkelsespunkt,

$$f_\infty(d_1(n), d_2(n)) = 20 \log_{10} \frac{\|d_1(n)\|_\infty}{\|d_2(n)\|_\infty}. \quad (2.28)$$

Det maksimale utslaget til et signal av endelig lengde N er gitt ved

$$m = \max_n |x(n)| = \|x(n)\|_\infty \quad (2.29)$$

Dette kriteriet måler hvor stor forskjell det er mellom det største utslaget i amplitude i fokuspunktet og det største utslaget i amplitude i ett av undertrykkelsespunktene.

Begge disse separasjonskriteriene måles i desibel(dB). Disse målene vil i oppgaven kalles *mål på den faktiske fokuseringen*. Når musikk eller tale skal fokuseres er det ikke opplagt om det relative maksimale utslaget, mellom fokuspunkt og et undertrykkelsespunkt, er et godt mål for å evaluere hvor god fokusering som oppnås. Empiriske lyttetester underveis i masteroppgaven, viser at dette forholdet ikke er et godt mål på hvor god fokusering som oppnås. Hvorfor bruker jeg da det relative maksimale utslaget mellom fokuspunkt og et undertrykkelsespunkt? Som vi skal se i kapittel 4 har god fokusering målt med maksimalt utslag en sammenheng med hvor godt lydfeltet som helhet lar seg kontrollere.

2.11.2 Separasjonsmål for kildeseparasjon

For kildeseparasjon brukes kriteriet SIR (*signal-to-interference-ratio*). Anta at vi har løst et kildeseparasjonsproblem, med en ønsket kilde $x_1(n)$, og to uønskede kilder $x_2(n)$ og $x_3(n)$, se seksjon 2.10.1. Energien til det ønskede $x_1(n)$ signal ved utgangen $y(n)$ er gitt ved $\sigma_{x_1}^2 \|d_1(n)\|$. Det ønskede signalet antas å være hvit støy med varians $\sigma_{x_1}^2$ [16]. Tilsvarende vil energien til de to uønskede signalene $x_2(n)$ og $x_3(n)$ ved utgangen $y(n)$ være $\sigma_{x_2}^2 \|d_2(n)\|$ og $\sigma_{x_3}^2 \|d_3(n)\|$.

$$\text{SIR}_{\text{ut}} = 10 \log_{10} \frac{\sigma_{x_1}^2 \|d_1(n)\|_2^2}{\sigma_{x_2}^2 \|d_2(n)\|_2^2 + \sigma_{x_3}^2 \|d_3(n)\|_2^2} \quad (2.30)$$

I [16] antas alle kildene å ha lik varians slik at SIR_{ut} forenkles til

$$\text{SIR}_{\text{ut}} = 10 \log_{10} \frac{\|d_1(n)\|_2^2}{\|d_2(n)\|_2^2 + \|d_3(n)\|_2^2} \quad (2.31)$$

Dette kriteriet kan tolkes som den relative styrken lyden til den ønskede kilden har, relativt til alle de uønskede lydkildene. Desto høyere denne verdien er desto bedre vil det ønskede signalet kunne høres, samtidig som de uønskede signalene vil svekkes.

I tillegg til separasjon/fokuserings-målene ser vi også på tilpasningsfeilen som også måles i dB.

$$\epsilon = 10 \log_{10} \|\mathbf{H}\mathbf{g} - \mathbf{d}\|_2^2. \quad (2.32)$$

2.12 Impulsresponsen i et rikt miljø

Hvordan skal vi oppnå vellykket fokusering/separasjon ved hjelp av metodene som her er beskrevet? Jeg legger teorien om impulsresponsen i et *rikt miljø* til grunn. Med et rikt miljø menes det at impulsresponsene vil ha tilstrekkelig forskjellig tidsdomene-representasjon i forskjellige retninger/punkter, og at de vil ha en innholdsrik tidsdomene-representasjon gitt av ekkoer.

I [36] anvender Sylvain Yon, Mickael Tanter og Mathias Fink tidsreverse-ring for å fokusere lyd med høyttalere i et rom. Dette rommet har gode refleksjonsegenskaper. Det vises at det oppnås bedre fokusering i et slikt rom, enn i ett simulert fritt felt. I oppfølger-artikkelen oppnås enda bedre resultater, da de fokuserer lyd med inversfiltrering[37]. De bedre resultatene skyldes blant annet at inversfiltrering kompenserer for tap av informasjon. Tapet av informasjon forekommer blant annet på grunn av absorpsjon av lyd i rommet. Innfallsvinkelen er her noe annerledes, da de løser det inverse problemet i frekvensdomenet. Forbedret fokusering er også observert innen telekommunikasjon i et uordnet medium[28].

Denne oppgaven er relativt nært knyttet til disse arbeidene, likheten ligger i tankegangen om at rikdom i miljøet bedrer fokuseringen. Tidligere arbeider undersøker større miljøer slik som rom. Jeg har i denne oppgaven valgt å begrense undersøkelsene til kun å betrakte et "lokalt rikt" miljø, hva som menes med dette vil bli nærmere forklart i neste kapittel.

I kapittel 4 skal jeg undersøke denne tankegangen med simulerte data. Dette følges opp i kapittel 5 da ekte data benyttes. De ekte dataene består av en database med hode-relaterte impulsresponsen, samt impulsresponsen målt i et rom. Ved disse impulsresponsmålingene er det gjort forsøk på å danne et lokalt rikt miljø ved å sette sylinder-aktige hindringer, plassert uordnet, foran høyttalerne.

2.13 Dårlig kondisjonerte likningsystemer

I praksis kan det oppstå mange overbestemte likningsystemer som gir opphav til et dårlig kondisjonert likningsystem. Dette kan være ved diskretisering av integrallikninger, inverse problemer, eller problemer innen signalbehandling[13]. En løsning av et dårlig kondisjonert system er sensitiv til perturbasjoner i både \mathbf{A} og \mathbf{b} .

De singulære verdiene til en matrise \mathbf{A} vil kunne gi innsikt om hvorvidt en matrise er dårlig kondisjonert. Man deler inn i to typer dårlig kondisjonerte

systemer[13]:

1. Problemer hvor singularverdiene til matrisen \mathbf{A} avtar gradvis mot null.
2. Problemer der distribusjonen av singularverdiene til \mathbf{A} har et veldefinert gap mellom de store og de veldig små singularverdiene.

Et annet mål på om en matrise er dårlig kondisjonert er, kondisjonstallet til en matrise \mathbf{A} . Dersom kondisjonstallet er veldig høyt er matrisen dårlig kondisjonert.

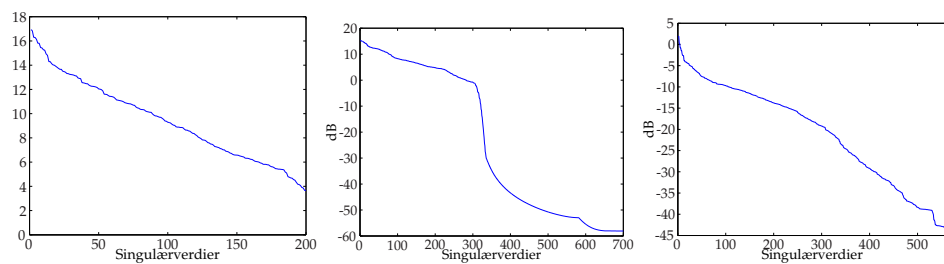
En løsning \mathbf{x} av et dårlig kondisjonert system kan gi veldig usikre resultater, om det i det hele tatt er mulig å frembringe en løsning[13]. Om man allikevel krever en løsning på slike problemer finnes det metoder som kan stabilisere løsningen. En slik stabilisering kalles regularisering.

Matrise sammensatt av konvolusjonsmatriser

La matrisen \mathbf{H} innholde konvolusjonsmatriser tilhørende simulerte data (se neste kapittel). \mathbf{H} er en matrise som fremkommet ved simulering av lydfokusering med to lydkilder, ett fokuspunkt og to undertrykkelsespunkter, impulsresponsene er av lengde 100. Matrisen vil være på formen som i likning 2.21. I figur 2.12(a) vises singularverdiene til denne matrisen. De singularverdiene til denne matrisen avtar ikke gradvis mot null, og de har heller ikke et veldefinert gap mellom store og små singularverdier. Kondisjonstallet til denne matrisen er 5, et lite kondisjonstall. Dette gir ikke implikasjoner på at denne matrisen er dårlig kondisjonert. De øvrige matrisene som inngår i simuleringene har tilsvarende singularverdikurver, og heller ikke høye kondisjonstall.

Foruten å jobbe med simulerte data, vil jeg i senere kapitler se på ekte, *målte data*, slik som hode-relaterte impulsresponser og impulsresponser målt gjennom et sett av reflektorer i et vanlig rom. Da vil vi blant annet oppleve at impulsresponsene som utgjør konvolusjonsmatrisene vil være båndbegrensede. Dette gir deretter en reduksjon i dimensjonen til kolonnerommet til matrisene fordi søylene blir "likere hverandre". Dette påvirker i sin tur fallet i singularverdier. Eksempler på dette er vist i figurene 2.12(b) og 2.12(c), og svarer til reelle situasjoner, henholdsvis med hode-relaterte impulsresponser og med målte impulsresponser gjennom reflektorer. Formen, og spesielt det bratte fallet på singularverdikurven, gir en indikasjon på at løsningen av de lineære systemene kan bli ustabile, spesielt hvis det er støy tilstede. Det finnes en rekke metoder å stabilisere slike løsninger

på, som f.eks ved å bruke regulariseringsmetoder slik som Tikhonov regularisering eller trunkert singulærverdi-dekomposisjon[13]. Å forfølge slike metoder er en relativt stor jobb, og vil ligge utenfor omfanget av denne oppgaven. Jeg antar derfor at jeg kan bruke normallikningene på standardform, og aksepterer at dette trolig vil gi dårligere estimater enn ved bruk av egnet regularisering. Vi merker oss imidlertid at de foreslåtte metodene sannsynligvis vil gi bedre resultater dersom de blir regularisert på en hensiktsmessig måte.



(a) Singulærverdispekt-
er tilhørende matrise \mathbf{H}
med konvolusjonsmat-
riser tilhørende simuler-
te impulsresponser.

(b) Singulærverdispekt-
er tilhørende matrise \mathbf{H}
med konvolusjonsmat-
riser tilhørende hode-
relaterte impulsrespon-
ser.

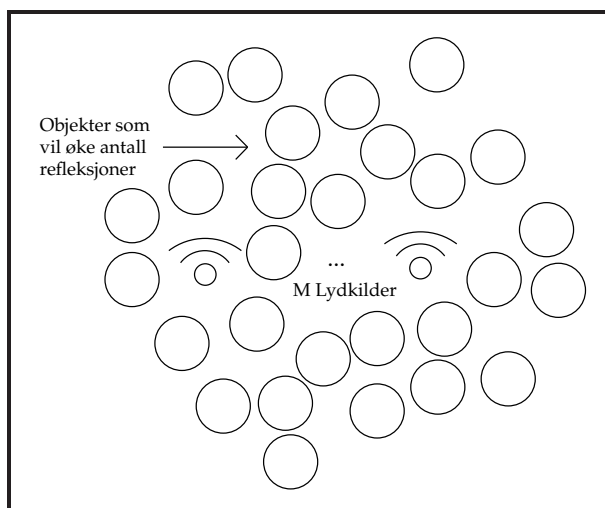
(c) Singulærverdispekt-
er tilhørende matrise \mathbf{H}
med konvolusjonsmat-
riser tilhørende impuls-
responser målt i et rom.

Figur 2.12

Kapittel 3

En simuleringsmodell

I dette kapitlet går jeg gjennom en modell laget for å utføre simuleringer til oppgaven. Modellen er implementert i MatlabTM. Vi har sett at kildeseparasjon og lydfokusering er nært beslektet, simulasjonsmodellen kan derfor betrakte begge problemstillingene. Simuleringene i neste kapittel vil kun betrakte lydfokuseringsproblemet, modellen vil derfor forklares i denne konteksten. Jeg antar at lydkildene er uni-direktive. Jeg antar videre at jeg kan gjøre lydutsendelsen direktiv ved å sette en konstellasjon av reflektorer rundt disse og kombinere dette med filtrering av utgangssignalet, se figur 3.1. Et slikt sett med reflektorer satt foran lydkildene vil heretter betegnes som et lokalt rikt miljø. Impulsresponsene som beskriver filterne mellom lydkilder og punkter målt et stykke fra reflektorene antas bare å inneholde ekkoe fra disse.



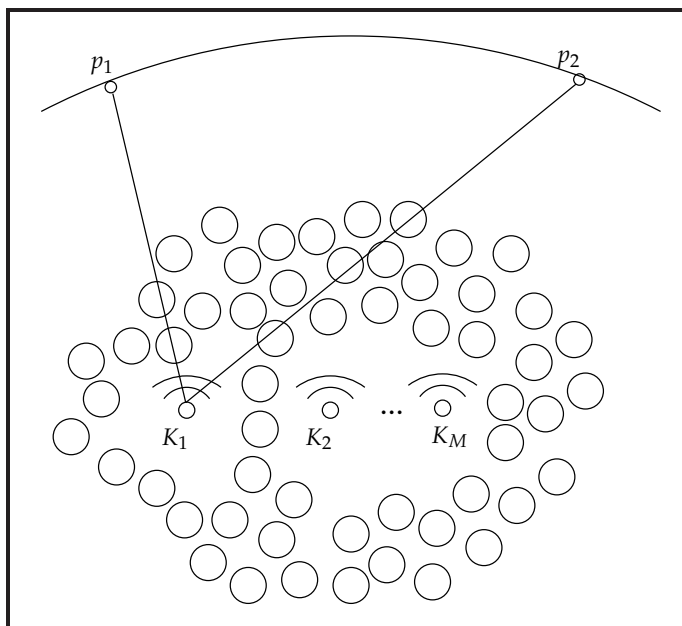
Figur 3.1: Reflektorer som vil øke spredningen av lydsignaler.

Betrakt figur 3.2, på grunn av reflektorene antar jeg at impulsresponsen fra lydkilden K_1 til punktet p_1 vil være forskjellig i forhold til impulsresponsen fra lydkilde K_1 til punktet p_2 .

For å modellere impulsresponsene mellom lydkilder og punkter trekker jeg uniformt distribuerte pseudo-tilfeldige tall i intervallet $[-1, 1]$, og gjør dem til filtertapper

$$h_{i,j}(n) \in [-1, 1]. \quad (3.1)$$

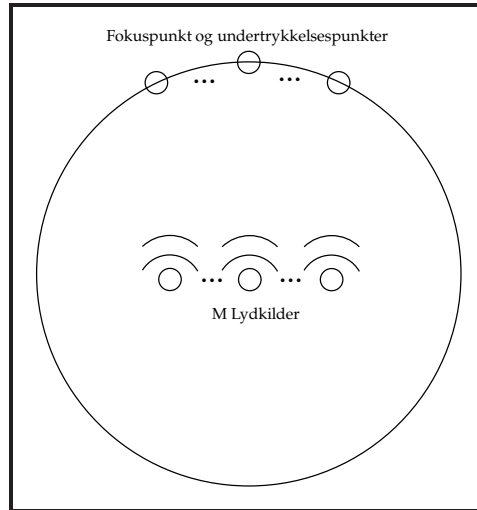
Antagelsen om en slik fordeling av amplitudebegrensede tilfeldige tapper virker rimelig ut fra intuisjon, og ut fra en tanke om at tilstrekkelig mange suksessive refleksjoner gir en diversitet som grenser mot det tilfeldige. Tilsvarende modeller benyttes også innenfor telekommunikasjon ved modellerering av kanaler signalene filtreres av mellom sendere og mottakere[10].



Figur 3.2: Reflektorer vil gi opphav til forskjellige impulsresponser i forskjellige punkter rundt lydkildene.

Jeg antar at fokuspunktet og de øvrige undertrykkelsespunktene er plassert på et kulegitter rundt lydkildene. Impulsresponser knyttes derfor til et punkt (ϕ, θ) , en horisontal vinkel ϕ og en elevasjonsvinkel θ . Antall impulsresponser som knyttes til et punkt avhenger av hvor mange lydkilder inne i kulegitteret som modelleres. Med M lydkilder knyttes det M

impulsresponser til hvert av punktene. Det er 1296 punkter på kulen. I figur 3.3 vises et utsnitt av modellen.

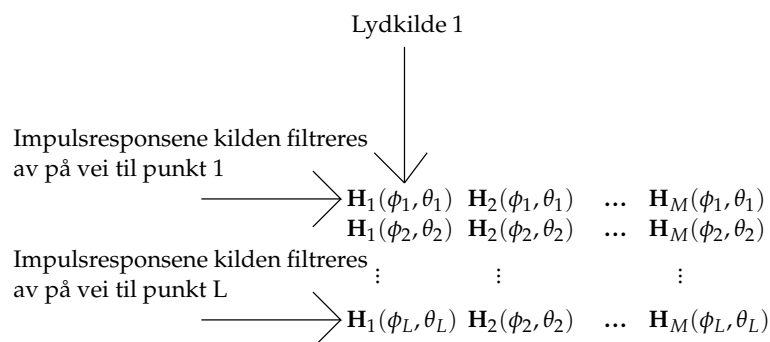


Figur 3.3: Utsnitt av simulasjonsmodellen ved lydfokusering, ingen elevasjon

3.1 Visualisering av fokuseringen

Anta at vi har et punkt $p_1 = (\phi_1, \theta_1)$ som vi ønsker å fokusere lydkildene inn mot. La punktene $p_2 = (\phi_2, \theta_2), \dots, p_L = (\phi_L, \theta_L)$ være punkter vi vil undertrykke. La filtermatrisene $\mathbf{H}_j(\phi_1, \theta_1)$, $j = 1, \dots, M$ være konvolusjonsmatrisene til filterne som beskriver systemene fra de M lydkildene til punktet p_1 . Disse utgjør øverste blokk-rad i den sammensatte konvolusjonsmatrisen \mathbf{H} i 2.22. Konvolusjonsmatrisene $\mathbf{H}_j(\phi_2, \theta_2)$, $j = 1, \dots, M$, som inneholder impulsresponsene fra lydkildene til p_2 utgjør andre blokk-rad i den sammensatte konvolusjonsmatrisen \mathbf{H} . På samme måte vil konvolusjonsmatrisene som inneholder impulsresponsene mellom lydkildene til punktene $p_3(\phi_3, \theta_3), \dots, p_L(\phi_L, \theta_L)$ danne nye blokk-rader i \mathbf{H} , se figur 3.4.

Løsningen av likningsystemet 2.22 vil gi filtre $g_1(n), \dots, g_M(n)$. Om lydkildene K_1, K_2, \dots, K_M prefiltreres av henholdsvis $g_1(n), g_2(n), \dots, g_M(n)$ vil lydkildene under ideelle forhold kunne fokusere et signal $x(n)$ inn mot punktet $p_1(\phi_1, \theta_1)$, og undertrykke $x(n)$ i de andre punktene $p_2(\phi_2, \theta_2), \dots, p_L(\phi_L, \theta_L)$. Etter endt inversfiltrering vil det i punktene være en respons



Figur 3.4: Hjelpesfigur til sammensetning av den sammensatte konvolusjonsmatrisen \mathbf{H} i modellen.

$$d(\phi_i, \theta_i, n) = \sum_{j=1}^M h_j(\phi_i, \theta_i, n) * g_j(n). \quad (3.2)$$

Denne responsen vil kunne gi informasjon om hvorvidt fokuseringen er vellykket. Responsen $d(\phi_i, \theta_i, n)$ brukes derfor til å plote resultatet av inversfiltreringen. I lys av kriteriene presentert i seksjon 2.11.1 velges plottkriteriene:

- Energi som slipper igjennom i et punkt på kulen $p_i = (\phi_i, \theta_i)$ etter endt inversfiltrering er

$$\begin{aligned} L_2(\phi_i, \theta_i) &= \|\mathbf{H}(\phi_i, \theta_i) \mathbf{g}\|_2^2 = \|[\mathbf{H}_1(\phi_i, \theta_i) \cdots \mathbf{H}_M(\phi_i, \theta_i)] \mathbf{g}\|_2^2 \\ &= \left\| \sum_{j=1}^M h_j(\phi_i, \theta_i, n) * g_j(n) \right\|_2^2. \end{aligned} \quad (3.3)$$

- Maksimalutslaget i et punkt $p_i(\phi_i, \theta_i)$ på kulen etter endt inversfiltrering er

$$\begin{aligned} L_\infty(\phi_i, \theta_i) &= \|\mathbf{H}(\phi, \theta) \mathbf{g}\|_\infty = \|[\mathbf{H}_1(\phi, \theta) \cdots \mathbf{H}_M(\phi, \theta)] \mathbf{g}\|_\infty \\ &= \left\| \sum_{j=1}^M h_j(\phi_i, \theta_i, n) * g(n) \right\|_\infty. \end{aligned} \quad (3.4)$$

I tillegg til å betrakte punktene som inngår i matriselikningen 2.22, undersøker jeg også hvordan filterne presterer i alle andre punkter på kulen, dvs hvor mye energi som måles i disse punktene. Filterne anvendes derfor på impulsresponsene tilhørende alle punkter. En algoritme for å visualisere fokuseringsprosessen er formulert som:

Algoritme for å visualisere fokuseringsprosessen.

1. Bestem fokuspunkt og undertrykkelsespunkter, og sett sammen matrisen \mathbf{H} . Bruk metodikken fra kapittel 2 til å konstruere filtre $g_1(n), g_2(n) \dots g_M(n)$
2. For hvert punkt på kulen, (ϕ_i, θ_i) , beregn normen $L_2(\phi_i, \theta_i)$ eller $L_\infty(\phi_i, \theta_i)$.
3. Plott verdiene av $L_2(\phi_i, \theta_i)$ eller $L_\infty(\phi_i, \theta_i)$ på tilhørende lapper/punkter på kulen.

3.2 Faktisk fokusering og utilsiktet undertrykking

Jeg innfører uttrykkene, *faktisk fokusering* og *utilsiktet undertrykking*. *Faktisk fokusering* er fokuseringen som inversfilterne er laget spesielt for, se seksjon 2.11.1. *Utilsiktet undertrykking* er undertrykkingen av alle andre punkter på kulen. Jeg innfører et mål for den utilsiktede undertrykkingen.

- Forholdet mellom energien som slippes igjennom i fokuspunktet og gjennomsnittet av energi som slippes igjennom i alle andre punkter enn de som inngår i matriselikningen 2.22.

$$f_2^- = 10 \log_{10} \frac{\|\mathbf{H}(\phi_{fp}, \theta_{fp}) \mathbf{g}\|_2^2}{\frac{1}{P} \left(\sum_r \sum_s \|\mathbf{H}(\theta_r, \phi_s) \mathbf{g}\|_2^2 \right)}. \quad (3.5)$$

Hvor r og s gjennomløper alle de vinklene som ikke er tatt med i den faktiske fokuseringen, og $p = rs$.

3.3 Romlig glatting av impulsresponsene

I simulasjonsmodellen vil det langs randen av kulen være punkter som har kort avstand til andre punkter. Kall nå to nærliggende punkter for $p_1 = (\phi_1, \theta_1)$ og $p_2 = (\phi_2, \theta_2)$, la et annet punkt $p_L = (\phi_L, \theta_L)$ være plassert et stykke unna p_1 og p_2 . Fra de M lydkildene til disse punktene, finnes det tre sett med impulsrespons, $h_j(\phi_1, \theta_1, n)$, $h_j(\phi_2, \theta_2, n)$ og $h_j(\phi_L, \theta_L, n)$ for $j = 1, \dots, M$. Disse tre settene med impulsrespons vil gi opphav til tre blokk-rader i den sammensatte konvolusjonsmatrisen,

$$\begin{bmatrix} \mathbf{H}_1(\phi_1, \theta_1) & \dots & \mathbf{H}_M(\phi_1, \theta_1) \\ \mathbf{H}_1(\phi_2, \theta_2) & \dots & \mathbf{H}_M(\phi_2, \theta_2) \\ \vdots & & \vdots \\ \mathbf{H}_1(\phi_L, \theta_L) & \dots & \mathbf{H}_M(\phi_L, \theta_L) \end{bmatrix}.$$

p_1 og p_2 ligger nær hverandre, jeg antar derfor at impulsresponsene som utgjør konvolusjonsmatrisene $\mathbf{H}_1(\phi_1, \theta_1)$ og $\mathbf{H}_1(\phi_2, \theta_2)$ vil være likere hverandre enn impulsresponsen som utgjør konvolusjonsmatrisen $\mathbf{H}_1(\phi_L, \theta_L)$, se figur 3.5. Dette gjelder også for parene av konvolusjonsmatriser $\mathbf{H}_j(\phi_1, \theta_1)$ og $\mathbf{H}_j(\phi_2, \theta_2)$, for $j = 2, \dots, M$.

Denne antagelsen begrunnes ved følgende argument: Refleksjoner fra lydkilder vil forplante seg relativt likt for målepunkter som ligger nær hverandre, mens for impulsresponsen som er målt et stykke fra hverandre på kulen, vil ekkoene forplate seg annerledes og impulsresponsene vil dermed få en mer ulik tidsdomene-representasjon.

Jeg har implementert en algoritme som glatter impulsresponsene i henhold til beskrivelsen over. Impulsresponsene knyttet til punktet $p(i, j)$ (jeg har byttet ut den horisontale vinkelen og elevasjonsvinkelen med henholdsvis i og j for lettere lesning) blir derfor en vektet sum av seg selv og impulsresponsene knyttet til nærliggende punkter. Til ethvert punkt på kulen kan det relateres en matrise med M impulsresponsen

$$\mathbf{P}(i, j) = \begin{bmatrix} h_1(0) & h_2(0) & \dots & h_M(0) \\ h_1(1) & h_2(1) & \dots & h_M(1) \\ \vdots & \vdots & \vdots & \vdots \\ h_1(N-1) & h_2(N-1) & \dots & h_M(N-1) \end{bmatrix} \quad (3.6)$$

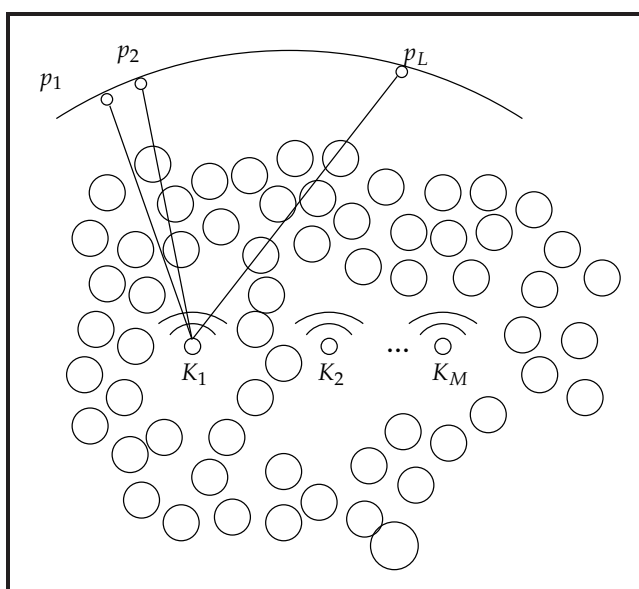
Jeg benytter meg av to forskjellige mønstre for glattingen, ett bestående av 9 punkter, og ett bestående av 25 punkter. Glattingen hvor impulsresponsene tilhørende punktet $p(i, j)$ er en vektet sum av seg selv og impulsresponsene tilhørende åtte nabopunkter beskrives ved

$$\mathbf{P}(i, j) = \frac{1}{9} \left(\mathbf{P}(i, j) + \sum_{k=-1}^1 \sum_{l=-1}^1 \mathbf{P}(i-k, j-l) \right). \quad (3.7)$$

Jeg kaller dette glattingsmønster 1. Glattingen hvor impulsresponsene tilhørende punktet $p(i, j)$ er en vektet sum av seg selv og impulsresponsene knyttet til tjuefire nabopunkter beskrives ved

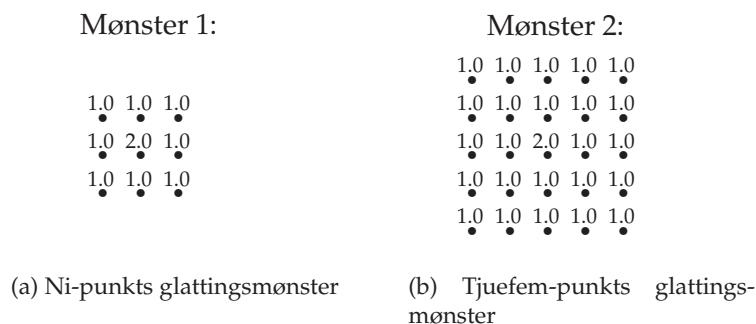
$$\mathbf{P}(i, j) = \frac{1}{25} \left(\mathbf{P}(i, j) + \sum_{k=-2}^2 \sum_{l=-2}^2 \mathbf{P}(i-k, j-l) \right). \quad (3.8)$$

Jeg kaller dette glattingsmønster 2. Disse to glattingsmønstre i punktform er vist i figur 3.6. Impulsresponsene knyttet til punktet i midten er de som blir en vektet sum av seg selv og impulsresponsene tilhørende nabopunktene. Glattingalgoritmen kan utføres flere ganger hvis ønskelig, for å gi økt total glatting. Nøyaktig hvor mye glatting som kan gjøres for å gjøre modellen mer lik virkeligheten er ikke opplagt. Vi må i utgangspunktet se på hvordan mengden glatting påvirker modellen og hva slags glatting som stemmer best overens med ekte, målte data.



Figur 3.5: Hjelpefigur for algoritmen som glatter impulsresponsene.

Det bør nevnes at jeg bare antar likhet mellom impulsresponser fra hver av de enkelte lydkildene til nærliggende punkter. Jeg antar dermed at det ikke oppstår noen likhet mellom impulsresponser fra to nærliggende lydkilder til et punkt. En slik likhet kan eventuelt forfølges i et senere arbeid.



Figur 3.6: To glattingsmønstre

Oppsummering av kriterier

- SIR_{ut} er et mål på hvor god separasjon som oppnås

$$SIR_{ut} = 10 \log_{10} \frac{\|d_1(n)\|_2^2}{\|d_2(n)\|_2^2 + \|d_3(n)\|_2^2}$$

- Relativt energinivå mellom fokuspunkt og et undertrykkelsespunkt

$$f_2(d_1(n), d_2(n)) = 10 \log_{10} \frac{\|d_1(n)\|_2^2}{\|d_2(n)\|_2^2}.$$

- Mål på den uttilsiktete undertrykkingen

$$f_2^- = 10 \log_{10} \frac{\|\mathbf{H}(\phi_{fp} \theta_{fp}) \mathbf{g}\|_2^2}{\frac{1}{P} \left(\sum_r \sum_s \|\mathbf{H}(\theta_r, \phi_s) \mathbf{g}\|_2^2 \right)}.$$

- Relativt maksimalt utslag mellom fokuspunkt og et undertrykkelsespunkt

$$f_\infty(d_1(n), d_2(n)) = 20 \log_{10} \frac{\|d_1(n)\|_\infty}{\|d_2(n)\|_\infty}.$$

- Plottkriteret som plotter hvor mye energi som slipper igjennom i et punkt etter endt inversfiltrering

$$\begin{aligned} L_2(\phi_i, \theta_i) &= \|\mathbf{H}(\phi_i, \theta_i) \mathbf{g}\|_2^2 = \|\left[\mathbf{H}_1(\phi_i, \theta_i) \cdots \mathbf{H}_M(\phi_i, \theta_i) \right] \mathbf{g}\|_2^2 \\ &= \left\| \sum_{j=1}^M h_j(\phi_i, \theta_i, n) * g_j(n) \right\|_2^2. \end{aligned}$$

- **Plottkriteret som plotter hvor utslag i amplitude som oppstår i ett punkt etter endt inversfiltrering**

$$\begin{aligned} L_{\infty}(\phi_i, \theta_i) &= \|\mathbf{H}(\phi, \theta) \mathbf{g}\|_{\infty} = \|[\mathbf{H}_1(\phi, \theta) \cdots \mathbf{H}_M(\phi, \theta) \mathbf{g}]\|_{\infty} \\ &= \left\| \sum_{j=1}^M h_j(\phi_i, \theta_i, n) * g(n) \right\|_{\infty}. \end{aligned} \quad (3.9)$$

Kapittel 4

Simuleringsresultater

Med simulasjonsmodellen fra forrige kapittel, er det utført en rekke forsøk. I dette kapitlet tar jeg for meg resultatene av disse simuleringene. Antagelsen om at økt rikdom i impulsresponsene vil øke fokuseringsegenskapene til systemet blir undersøkt, og til dels styrket av disse forsøkene. Jeg har funnet ut at dette gjelder både i tid og rom. Rikdom i tid og at impulsreponser er forskjellige fra hverandre i rom spiller inn på forskjellige fokuseringsmål; Tilsammen spiller de inn på hvor godt lydfeltet kan kontrolleres. Forsøkene tar også for seg forskjellige antall lydkilder og undertrykkelsespunkter.

4.1 Faktorer som undersøkes

Mengden situasjoner som kan oppstå når man har 1296 mulige fokuspunkter på kulen, med vilkårlige mange lydkilder og undertrykkelsespunkter, er enormt. Det blir følgelig umulig å dekke alle. Videre har jeg konsentrert forsøkene om noen utvalgte faktorer som jeg mener er viktige. De faktorene som undersøkes er:

1. Hvordan mengden glatting påvirker fokuseringsegenskapene til systemet under den foreslåtte modellen.
2. Hvorvidt innholdsrike impulsresponsen vil bidra til økte fokuseringsegenskaper.
3. Hvordan inverse filtere med lengder større enn de tilhørende impulsresponsene påvirker de globale fokuseringsegenskapene.
4. Hvordan forholdet mellom antall lydkilder og antall undertrykkelsespunkter påvirker fokuseringen. .
5. Plassering av undertrykkelsespunktene i forhold til fokuspunkt og mengden glatting.

4.2 Mengden glatting og innvirkning på fokuseringsegenskapene til modellen

Innledningsvis skal vi se hvordan mengden glatting påvirker den faktiske fokuseringen. Simuleringer viser at likhet mellom impulsresponsene, som utgjør konvolusjonsmatrisene langs blokk-kolonnene i den sammensatte konvolusjonsmatrisen \mathbf{H} , gjør den faktiske fokuseringen målt med relative energinivåer dårligere, enn når impulsresponsene har mer ulike tidsdomene-representasjon. Dette vises ved å løse inverse problemer, der impulsresponsene hentes fra simulasjonsmodeller, hvor ulike mengde glatting er påført.

Jeg tar utgangspunkt i en modell med to lydkilder, ett fokuspunkt, og to undertrykkelsespunkter. Jeg utfører simuleringer der jeg varierer mengden glatting som påføres modellen. Fire forskjellige glattingsskjemaer betraktes:

- Ingen glatting.
- Glattingsskjema 1: Glattingensmønster 1, algoritmen utført 1 gang.
- Glattingsskjema 2: Glattingensmønster 1, algoritmen utført 3 ganger.
- Glattingsskjema 3: Glattingensmønster 2, algoritmen utført 6 ganger.

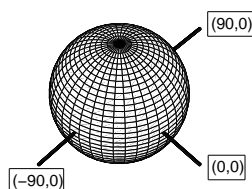
Jeg utfører glattingsalgoritmen flere ganger for å oppnå økt total glatting. Ved å utføre glattingsalgoritmen flere ganger resulterer dette også i at likheten mellom impulsresponsene økes ytterligere (jamfør 3.3).

Fokuspunktet settes til å være fast i $(\phi, \theta) = (0, 0)$. Undertrykkelsespunktene varieres i posisjonene

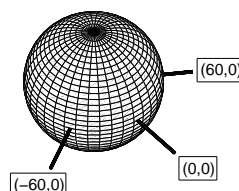
- (i) $(-90, 0)$ og $(90, 0)$, se figur 4.1(a)
- (ii) $(-60, 0)$ og $(60, 0)$, se figur 4.1(b)
- (iii) $(-30, 0)$ og $(30, 0)$, se figur 4.1(c)
- (iv) $(-10, 0)$ og $(10, 0)$, se figur 4.1(d)

Jeg modellerer et "lokalt miljø", og antar derfor relativt korte impulsresponser, med 100 tapper. Dette tilsvarer ca. to millisekunder med en sample-rate på 44,1KHz. Jeg betrakter først gjennomsnittlige relative energinivåer, fra 100 modell-realiseringer.

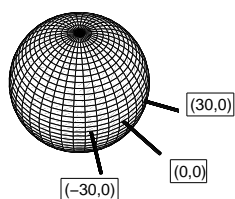
Tabell 4.1 viser fire undertabeller. De fire undertabellene inneholder gjennomsnittlige relative energinivåer, $f_2(.,.)$, mellom fokuspunktet og de to undertrykkelsespunktene, fra modeller som er påført ulike mengde glatting.



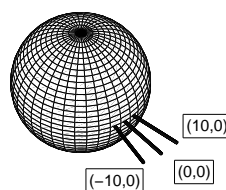
(a) Plasseringer av fokuspunkt og undertrykkelsespunkter til oppsett (i)



(b) Plasseringer av fokuspunkt og undertrykkelsespunkter til oppsett (ii)



(c) Plasseringer av fokuspunkt og undertrykkelsespunkter til oppsett (iii)



(d) Plasseringer av fokuspunkt og undertrykkelsespunkter til oppsett (iv)

Figur 4.1: Posisjonene til fokuspunkt og undertrykkelsespunkter tilhørende oppsettene (i), (ii), (iii) og (iv)

Tabell 4.1(c) inneholder gjennomsnittlige relative energinivåer fra modeller der glattingsskjema to er benyttet. Her ser vi at etterhvert som undertrykkelsespunktene flyttes nærmere fokuspunktet, avtar de relative energinivåene. Dette ser vi i større grad i tabell 4.1(d), som inneholder gjennomsnittlige relative energinivåer fra modeller der glattingsskjema tre er påført. Her inntreffer det et tap i $f_2(\cdot, \cdot)$ med undertrykkelsespunkter plassert ennå lengre vekk fra fokuspunktet. De gjennomsnittlige relative energinivåene fra modeller der glattingsskjema én påføres, avtar først når undertrykkelsespunktene er plassert ti grader fra fokuspunktet. De avtar også mindre enn for relative energinivåer fra modeller påført mer glatting. Dersom ingen glatting påføres, er de relative energinivåene så og si uforandret uansett plassering.

| | Grader | $f_2(.,.)$ |
|-------|------------|------------|
| (i) | -90 grader | 6.2dB |
| | 90 grader | 6.3dB |
| (ii) | -60 grader | 6.2dB |
| | 60 grader | 6.2dB |
| (iii) | -30 grader | 6.2dB |
| | 30 grader | 6.2dB |
| (iv) | -10 grader | 6.2dB |
| | 10 grader | 6.2dB |

(a) Ingen glatting

| | Grader | $f_2(.,.)$ |
|-------|------------|------------|
| (i) | -90 grader | 6.2dB |
| | 90 grader | 6.3dB |
| (ii) | -60 grader | 6.2dB |
| | 60 grader | 6.3dB |
| (iii) | -30 grader | 6.2dB |
| | 30 grader | 6.3dB |
| (iv) | -10 grader | 4.0dB |
| | 10 grader | 4.0dB |

(b) Glattingsmønster 1, algoritmen utført 1 gang

| | Grader | $f_2(.,.)$ |
|-------|------------|------------|
| (i) | -90 grader | 6.2dB |
| | 90 grader | 6.2dB |
| (ii) | -60 grader | 6.2dB |
| | 60 grader | 6.2dB |
| (iii) | -30 grader | 5.7dB |
| | 30 grader | 5.8dB |
| (iv) | -10 grader | 1.2dB |
| | 10 grader | 1.2dB |

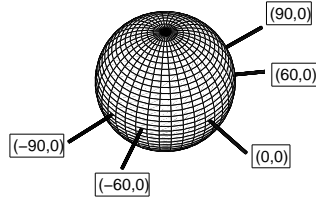
(c) Glattingsmønster 1, algoritmen utført 3 ganger

| | Grader | $f_2(.,.)$ |
|-------|------------|------------|
| (i) | -90 grader | 6.1dB |
| | 90 grader | 6.1dB |
| (ii) | -60 grader | 5.1dB |
| | 60 grader | 5.2dB |
| (iii) | -30 grader | 1.9dB |
| | 30 grader | 1.9dB |
| (iv) | -10 grader | 0.4dB |
| | 10 grader | 0.2dB |

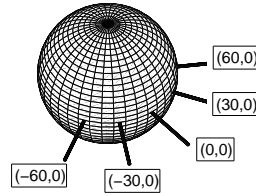
(d) Glattingsmønster 2, algoritmen utført 6 ganger

Tabell 4.1: Gjennomsnittlige relative energinivåer, mellom fokuspunkt og undertrykkelsespunkter, fra modeller med to lydkilder, ett fokuspunkt, to undertrykkelsespunkter og forskjellig mengder glatting

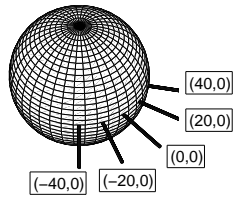
Det er kjent fra antenneteori at en økning av sensorer øker oppløsningen til antennen[19]. En økning i antall lydkilder bør også her gi bedre resultater. Vi betrakter derfor en modell med fire lydkilder, ett fokuspunkt og fire undertrykkelsespunkter. Glattingsskjemaene lar vi være like som tidligere i dette delkapitlet, og vi lar fokuspunktet være fast i $(\phi, \theta) = (0, 0)$. Fire forskjellige plasseringer av undertrykkelsespunkter betraktes.



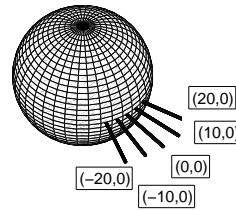
(a) Plasseringer av fokuspunkt og undertrykkelsespunkter til oppsett (v)



(b) Plasseringer av fokuspunkt og undertrykkelsespunkter til oppsett (vi)



(c) Plasseringer av fokuspunkt og undertrykkelsespunkter til oppsett (vii)



(d) Plasseringer av fokuspunkt og undertrykkelsespunkter til oppsett (viii)

Figur 4.2: Posisjonene til fokuspunkt og undertrykkelsespunkter tilhørende oppsettene (v), (vi), (vii) og (viii).

- (v) $(-90,0)$, $(90,0)$, $(-60,0)$ og $(60,0)$, se figur 4.2(a).
- (vi) $(-60,0)$, $(60,0)$, $(-30,0)$ og $(30,0)$, se figur 4.2(b).
- (vii) $(-40,0)$, $(40,0)$, $(-20,0)$ og $(20,0)$, se figur 4.2(c).
- (viii) $(-20,0)$, $(20,0)$, $(-10,0)$ og $(10,0)$, se figur 4.2(d).

Tabell 4.2 viser fire undertabeller med gjennomsnittlige relative energinivåer, mellom fokuspunkt og undertrykkelsespunktene, fra modeller påført de fire ulike glattingsskjemaene. De relative energinivåene er også her gjennomsnittet fra 100 modell-realiseringer. Vi legger først merke til at verdiene generelt er høyere, noe som er forventet siden vi har økt antall lydkilder.

| | Grader | $f_2(.,.)$ |
|--------|------------|------------|
| (v) | -90 grader | 9.5dB |
| | 90 grader | 9.5dB |
| | -60 grader | 9.5dB |
| | 60 grader | 9.6dB |
| (vi) | -60 grader | 9.5dB |
| | 60 grader | 9.5dB |
| | -30 grader | 9.5dB |
| | 30 grader | 9.5dB |
| (vii) | -40 grader | 9.5dB |
| | 40 grader | 9.5dB |
| | -20 grader | 9.5dB |
| | 20 grader | 9.5dB |
| (viii) | -20 grader | 9.6dB |
| | 20 grader | 9.5dB |
| | -10 grader | 9.6dB |
| | 10 grader | 9.5dB |

(a) Ingen glatting

| | Grader | $f_2(.,.)$ |
|--------|------------|------------|
| (v) | -90 grader | 9.6dB |
| | 90 grader | 9.7dB |
| | -60 grader | 9.5dB |
| | 60 grader | 9.5dB |
| (vi) | -60 grader | 9.5dB |
| | 60 grader | 9.5dB |
| | -30 grader | 9.5dB |
| | 30 grader | 9.5dB |
| (vii) | -40 grader | 9.4dB |
| | 40 grader | 9.5dB |
| | -20 grader | 9.2dB |
| | 20 grader | 9.3dB |
| (viii) | -20 grader | 9.2dB |
| | 20 grader | 9.3dB |
| | -10 grader | 6.7dB |
| | 10 grader | 6.8dB |

(b) Glattingsmønster 1, algoritmen utført 1 gang

| | Grader | $f_2(.,.)$ |
|--------|------------|------------|
| (v) | -90 grader | 9.7dB |
| | 90 grader | 9.7dB |
| | -60 grader | 9.7dB |
| | 60 grader | 9.7dB |
| (vi) | -60 grader | 9.5dB |
| | 60 grader | 9.5dB |
| | -30 grader | 9.1dB |
| | 30 grader | 9.0dB |
| (vii) | -40 grader | 9.2dB |
| | 40 grader | 9.2dB |
| | -20 grader | 6.7dB |
| | 20 grader | 6.7dB |
| (viii) | -20 grader | 9.5dB |
| | 20 grader | 9.5dB |
| | -10 grader | 3.2dB |
| | 10 grader | 3.2dB |

(c) Glattingsmønster 1, algoritmen utført 3 gang

| | Grader | $f_2(.,.)$ |
|--------|------------|------------|
| (v) | -90 grader | 10.7dB |
| | 90 grader | 10.8dB |
| | -60 grader | 9.4dB |
| | 60 grader | 9.4dB |
| (vi) | -60 grader | 9.8dB |
| | 60 grader | 9.8dB |
| | -30 grader | 3.8dB |
| | 30 grader | 3.8dB |
| (vii) | -40 grader | 8.9dB |
| | 40 grader | 8.8dB |
| | -20 grader | 2.8dB |
| | 20 grader | 2.8dB |
| (viii) | -20 grader | 5.8dB |
| | 20 grader | 5.9dB |
| | -10 grader | 1.9dB |
| | 10 grader | 1.9dB |

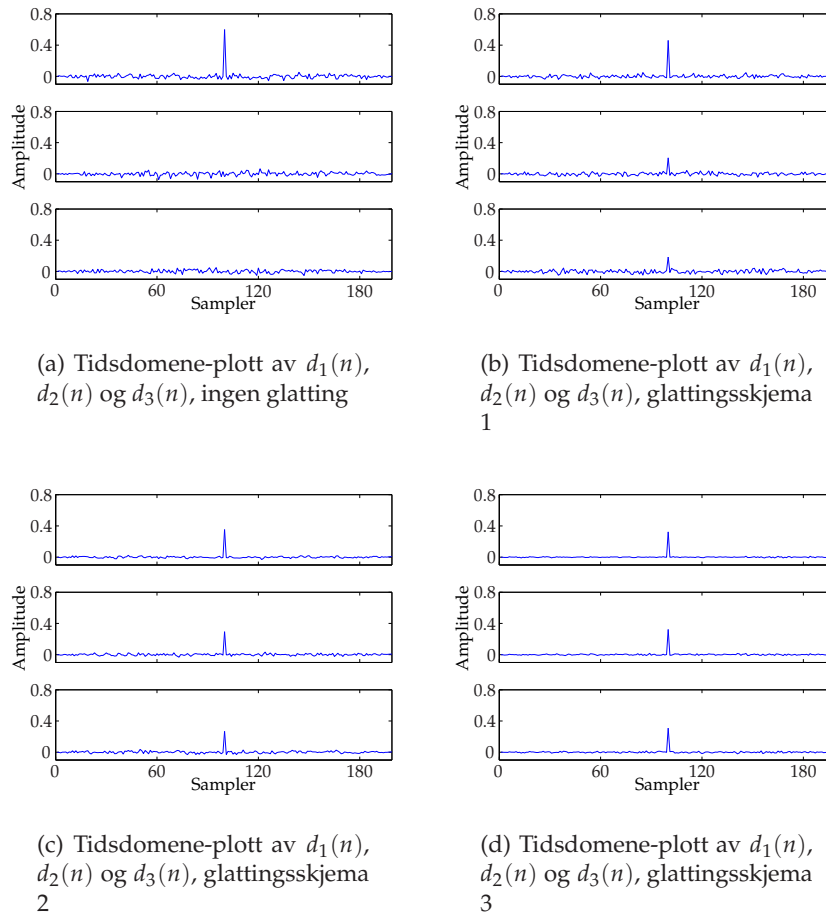
(d) Glattingsmønster 2, algoritmen utført 6 ganger

Tabell 4.2: Gjennomsnittlige relative energinivåer, mellom fokuspunkt og undertrykkelsespunkter, fra modeller med 4 lydkilder, ett fokuspunkt, 4 undertrykkelsespunkter og forskjellig mengder glatting

Tabellene viser de samme tendensene som ved forrige modell, jo mere glatting som påføres, jo lengre vekk fra fokus synker de relative energinivåene. Uten glatting forblir verdiene de samme.

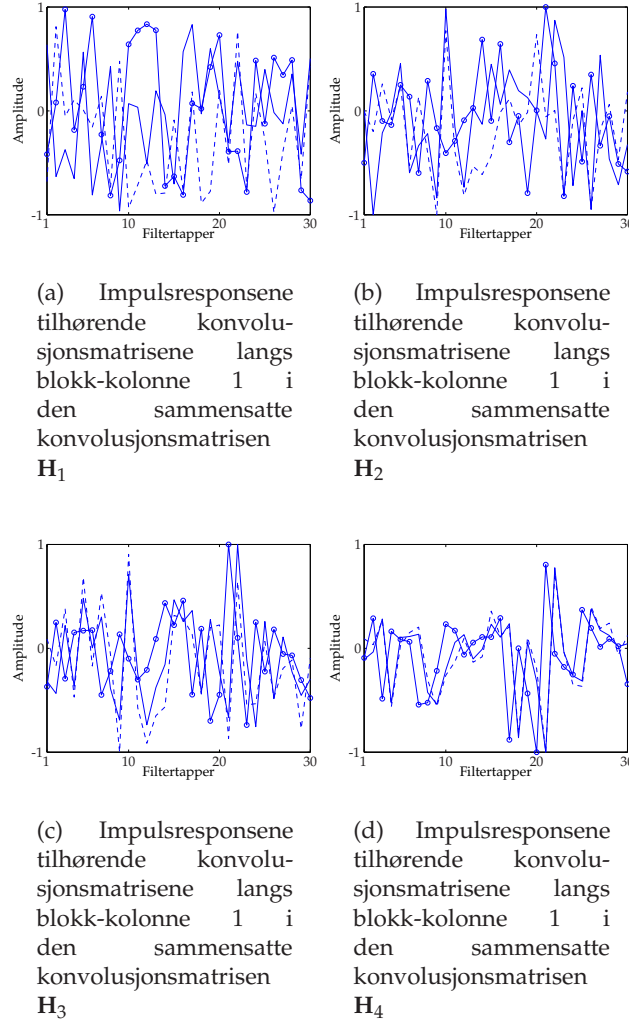
4.2.1 Mengden glatting og innvirkning på den estimerte responsen

Videre ser vi på tidsdomene-plott av den estimerte responsen $\mathbf{H}\mathbf{g} = \hat{\mathbf{d}}$, der \mathbf{g} er løsningen av likningsystemet 2.21. Med ett fokuspunkt og to undertrykkelsespunkter består den estimerte responsvektoren $\hat{\mathbf{d}}$ av tre delvektorer. I det ideelle tilfellet skal øverste delvektor inneholde de diskrete verdiene av en delta-funksjon, mens de to øvrige skal kun inneholde nuller. La \mathbf{d}_1 være øverste delvektor av $\hat{\mathbf{d}}$ og \mathbf{d}_2 og \mathbf{d}_3 være henholdsvis andre og tredje delvektor av $\hat{\mathbf{d}}$.



Figur 4.3: Plott av responsen i fokuspunktet og undertrykkelsespunktene ved forskjellige mengder glatting

I figur 4.3 er delvektorene til fire estimerte responser $\hat{\mathbf{d}}$ plottet. Responsene er hentet fra fire modeller med de samme glattingsskjemaene brukt



Figur 4.4: Impulsresponser fra kolonnene til de sammensatte konvolusjonsmatrisene \mathbf{H}_1 , \mathbf{H}_2 , \mathbf{H}_3 og \mathbf{H}_4

tidligere i dette delkapitlet. Undertrykkelsespunktene er plassert ti grader fra fokuspunktet, se figur 4.1(d). Enkelt-realiseringer av modellen er brukt her. I plott 4.3(b) (glattingsskjema 1) er det synlig en liten topp i responsen til undertrykkelsespunktene. Disse toppene er ikke tilstede i plott 4.3(a) (ingen glatting). Etterhvert som mer glatting påføres modellen blir toppene mere distinkte i responsene. I figur 4.3(d) har responsen til undertrykkelsespunktene lik form som responsen til fokuspunktet. Her forekommer det ingen fokusering. De relative energinivåene for disse plottene er (se seksjon 2.11.1):

- For figur 4.3(a): $f_2(d_1(n), d_2(n)) = 7dB$ $f_2(d_1(n), d_3(n)) = 8.2dB$
- For figur 4.3(b): $f_2(d_1(n), d_2(n)) = 4.4dB$ $f_2(d_1(n), d_3(n)) = 3.9dB$
- For figur 4.3(c): $f_2(d_1(n), d_2(n)) = 1dB$ $f_2(d_1(n), d_3(n)) = 1.5dB$
- For figur 4.3(d): $f_2(d_1(n), d_2(n)) = -0.1dB$ $f_2(d_1(n), d_3(n)) = 0.2dB$

4.2.2 Mengden glatting og likhet mellom impulsresponser

Mer glatting resulterer i dårligere faktisk fokusering. Vi skal nå se på impulsresponsene som inngår i det inverse problemet, ved forskjellig mengde glatting.

Kall nå den sammensatte konvolusjonsmatrisen fra modellen uten glatting for \mathbf{H}_1 , og kall matrisen fra modellen påført glattingsskjema én for \mathbf{H}_2 . Den sammensatte konvolusjonsmatrisen fra modellen påført glattingsskjema to, kaller vi \mathbf{H}_3 , og den som fremkommer ved at glattingsskjema tre påføres modellen, kaller vi \mathbf{H}_4 .

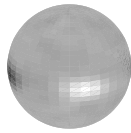
I figur 4.4(a) plottes 30 tapper av tre impulsresponser. Impulsresponsene utgjør konvolusjonsmatrisene som ligger langs første blokk-kolonne i \mathbf{H}_1 . 30 tapper av de tre impulsresponsene som utgjør konvolusjonsmatrisene langs blokk-kolonnene i \mathbf{H}_2 , \mathbf{H}_3 og \mathbf{H}_4 , er plottet i henholdsvis i figurene 4.4(b), 4.4(c) og 4.4(d).

I figur 4.4(a) er det ingen tilsynelatende likhet mellom impulsresponsene. Dette gjelder forøvrig også i figur 4.4(b), men vi kan se en viss likhet i formen. Impulsresponsene i figurene 4.4(c) og 4.4(d) har fått større likhet mellom seg. De er fremdeles ikke helt like, men de følger stort sett de samme trendene.

Vi ser her at impulsresponser med mer lik tidsdomene-representasjonene, fra hver av lydkildene til punkter på kulen, gjør den faktiske fokuseringen dårligere.

4.2.3 Hvordan mengden glatting påvirker den globale fokuseringen

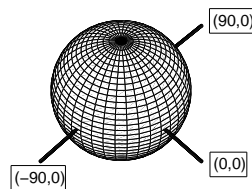
Vi skal til slutt i dette delkapitlet se på plott av L_2 -verdier og hvordan mengden glatting har innvirkning på størrelsen på fokuspunktet. Plottene viser L_2 -verdier beregnet fra modeller med to lydkilder, ett fokuspunkt og to undertrykkelsespunkter. Plott av L_2 -verdier beregnet fra modeller med fire lydkilder, ett fokuspunkt og fire undertrykkelsespunkter vises også. Modellene er påført glattingsskjema to og tre. Vi betrakter et gjennomsnitt



(a) Plott av gjennomsnittlige L_2 -verdier, glattingsmønster 1, algoritmen utført 3 ganger.



(b) Plott av gjennomsnittlige L_2 -verdiene, glattingsmønster 2, algoritmen utført 6 ganger.



(c) Hjelpesfigur til plassering av fokuspunkt og undertrykkelsespunkter. Sylinderne merket "Fokuspunkt" skjærer kulen i fokuspunktet.

| | |
|-----|-------|
| (a) | 1.9dB |
| (b) | 2.0dB |

Figur 4.5: Plott av gjennomsnittlige L_2 -verdier fra modeller med 2 lydkilder, 2 undertrykkelsespunkter, og ett fokuspunkt. Glattingsmønstre, 1(a) og 2(b), algoritmen utført henholdsvis 3(a) og 6(b) ganger. Med de tilhørende gjennomsnittlige utilsiktede relative energinivåene.

av 100 modell-realiseringer. Dette betyr at L_2 -verdiene er beregnet 100 ganger og gjennomsnittet av disse er plottet. Gjennomsnittlige L_2 -verdier beregnet fra modeller med to lydkilder kan sees i figur 4.5. Det lyse området på framsiden av kulene er fokuspunktet. Det mørke området som er plassert 90 grader til siden for fokuspunktet er det ene undertrykkelsespunktet. Vi ser at det oppnås en viss utilsiktet undertrykking. Verdiene på den util-

siktede undertrykkingen kan sees i den tilhørende tabellen i figur 4.5. Disse verdiene er lavere enn de relative energinivåene mellom fokuspunkt og undertrykkelsespunkter, se tabell 4.1, men det er interessant at det i det hele tatt forekommer.

Ved sammenlikning av plottene merker vi oss at fokuspunktet i figur 4.5(a) er mindre enn for det tilsvarende området i figur 4.5(b). Dette er også tilfellet for undertrykkelsespunktene. Her kan glattingen ha en positiv effekt ved at et større område undertrykkes.

For modellen med fire lydkilder, ser vi at situasjonen er den samme, se figur 4.6. Fokuspunktet utvides ettersom mer glatting påføres. Fra tabellene 4.1 og 4.2 vet vi at den faktiske fokuseringen bedres med en økning fra to til fire lydkilder. Vi legger merke til at den utilsiktede undertrykkingen også bedres.

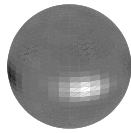
Kommentar til dette delkapitlet

I dette delkapitlet har vi sett på hvordan mengden glatting som påføres påvirker fokuseringsegenskapene til systemet under den foreslåtte modellen. Simuleringene har vist at glattingen påvirker den faktiske fokuseringen, og at den har en innvirkning på størrelsen på fokuspunktet. Det blir dermed viktig at impulsresponser har ulik tidsdomene-representasjon, fra hver av lydkildene til punkter på kulen, for å oppnå god fokusering.

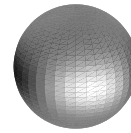
4.3 Lengden på impulsresponsene og innvirkningen på fokuseringsegenskapene

Som fortalt i seksjonene 2.12 og 1.1.3 har tidligere forskning vist at om man forsøker å fokusere lyd med tidsreversering i et rom som ikke er lyddempet, kan fokuseringsegenskapene til systemet bedres. Ekkoene fra vegger, tak, gulv vil resultere i en virtuell høyttalerantenne som er større enn den faktiske. Ekkoene vil virtuell utvide antennen, slik at man i praksis vil få en større antenne enn den som fremstår i virkeligheten.

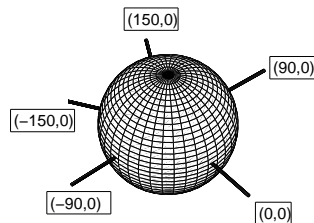
Denne tankegangen kan forfølges ved å undersøke hvordan lengre impulsresponser påvirker fokuseringsegenskapene til modellen. Vi betrakter fire modeller, alle med to lydkilder, ett fokuspunkt, og to undertrykkelsespunkter. Fokuspunktet er plassert i $(0,0)$, og undertrykkelsespunktene er plassert i $(-90,0)$ og $(90,0)$. Responsen i fokuspunktet er $d_1(n)$, og responsen i undertrykkelsespunktene er $d_2(n)$, og $d_3(n)$. De fire modellene har impulsresponslengder 10, 100, 200, og 500. Jeg antar at "rikdommen i ekkoer",



(a) Plott av gjennomsnittlige L_2 -verdier, glattingsmønster 1, algoritmen utført 3 ganger.



(b) Plott av gjennomsnittlige L_2 -verdier, glattingsmønster 2, algoritmen utført 6 ganger.



(c) Hjelpesfigur til plassering av fokuspunkt og undertrykkelsespunkter.

| | |
|-----|-------|
| (a) | 3.6dB |
| (b) | 3.5dB |

Figur 4.6: Plott av gjennomsnittlige L_2 -verdier fra modeller med 4 lydkilder, 4 undertrykkelsespunkter og ett fokuspunkt. Glattingsmønstre, 1(a) og 2(b), algoritmen utført henholdsvis 3(a) og 6(b) ganger. Med de tilhørende gjennomsnittlige utilsiktede relative energinivåene.

gitt ved lengden av impulsresponsene kan gjøres større om jeg øker antallet objekter som øker spredningen av signalet. Den faktiske fokuseringen undersøkes først. Denne måles med både relative energinivåer, $f_2(.,.)$, og relative maksimale utslag, $f_\infty(.,.)$. Tilpasningsfeilen ϵ , knyttet til det inverse problemet, undersøkes også. Her er ingen glatting påført modellen.

| IR-lengde | $f_2(d_1(n), d_2(n))$ | $f_2(d_1(n), d_3(n))$ | $f_\infty(d_1(n), d_2(n))$ | $f_\infty(d_1(n), d_3(n))$ | ϵ |
|-----------|-----------------------|-----------------------|----------------------------|----------------------------|------------|
| 10 | 6.2dB | 6.3dB | 12.6dB | 12.6dB | -3.4dB |
| 100 | 6.2dB | 6.2dB | 19.0dB | 19.0dB | -3.4dB |
| 200 | 6.2dB | 6.2dB | 21.2dB | 21.2dB | -3.4dB |
| 500 | 6.2dB | 6.2dB | 24.4dB | 24.4dB | -3.4dB |

Tabell 4.3: Gjennomsnittlige relative energinivåer, relative maksimale utslag og tilpasningsfeil fra modeller med impulsresponser av lengde 10, 100, 200 og 500

4.3.1 Gjennomsnittlige relative energinivåer, relative maksimale utslag og impulsresponser med forskjellig mengder rikdom.

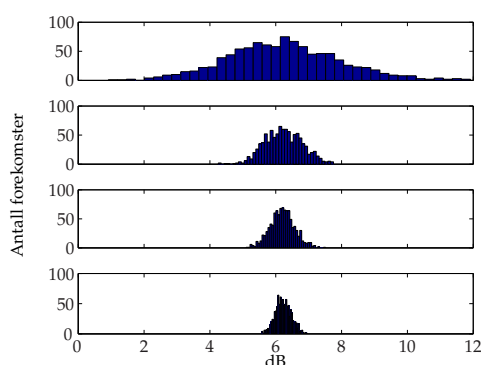
Tabell 4.3 inneholder gjennomsnittlige relative energinivåer, relative maksimale utslag, samt tilpasningsfeilen tilknyttet det inverse problemet fra modeller med impulsresponser av lengde 10, 100, 200 og 500. Verdiene er gjennomsnittet av verdier fra 1000 realiseringer av modellen.

Det første vi legger merke til i tabell 4.3, er at de relative energinivåene og tilpasningsfeilen forblir uendret ettersom lengden på impulsresponsene øker. Ved bruk av større matriser og vektorer vil flere komponenter kunne innholde "feil". Ved bruk av lengre impulsresponser vil "feilen" spres utover flere mulige komponenter i den estimerte responsen $\hat{\mathbf{d}}$.

De verdiene som bedres med økte impulsresponslengder er de relative maksimale utslagene. Nærmest en fordobling forekommer i modeller med impulsresponslengde 500, i forhold til de med impulsresponser av lengde 10. Dette betyr at den maksimale amplituden til responsen til undertrykkelsespunktene presses ned ved at lengre impulsresponser inngår i det inverse problemet.

De relative energinivåene, mellom fokuspunktet og undertrykkelsespunktene, er det kriteriet som best forklarer fokuseringen om systemet er tenkt benyttet til applikasjoner for hørbar lyd. Betyr dette at lengre impulsresponser ikke bedrer fokuseringsegenskapene til modellen? Dette virker motstridende i forhold til annen relatert forskning, og undersøkes nærmere nedenfor.

På bakgrunn av dette undersøker jeg fordelingene av de gjennomsnittlige relative energinivåene, de gjennomsnittlige relative maksimale utslagene og tilpasningsfeilen fra modellene med forskjellige impulsresponslengder. Disse fordelingene vises henholdsvis i figurene 4.7, 4.8 og 4.9.



Figur 4.7: Fordeling av relative energinivåer ved forskjellige impulsresponslengder. Den øverste fordelingen er fra modeller med impulsresponslengde 10, den nest øverste fra modeller med impulsresponslengde 100, den tredje fra modeller med impulsresponslengde 200 og den siste fra modeller med impulsresponslengde 500.

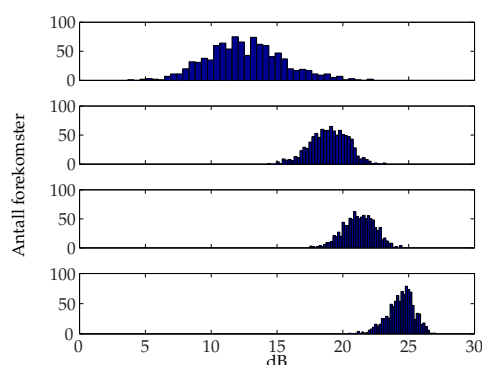
Et fellestrekk ved fordelingene er at etterhvert som lengden til impulsresponsen økes, konsentreres fordelingene om gjennomsnittsverdiene, vist i tabell 4.3. For impulsresponser med lengde 10 varierer fordelingene (de øverste fordelingene i figurene 4.7, 4.8 og 4.9) i stor grad. Det kan forekomme situasjoner hvor man oppnår bedre fokusering enn med lengre impulsresponser. Dette gjelder også i negativ retning, fokuseringen kan også bli dårligere. Det samme gjelder for fordelingene til tilpasningsfeilen.

4.3.2 Impulsresponslengder og globale fokuseringsegenskaper ved modellen

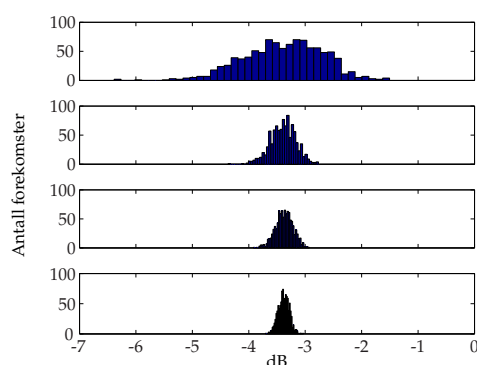
Videre undersøkes det hvordan innholdsrike impulsresponser spiller inn på hvor godt lydfeltet på kulen kan kontrolleres. Fire forskjellige modeller betraktes, alle med et fokuspunkt i $(0,0)$, og to undertrykkelsespunkter, plassert i punktene $(-90,0)$ og $(90,0)$. Modellene har impulsresponser av lengde 10, 100, 200 og 500, og de er påført glattingsskjema 1, se delkapittel 4.2.

Fra modellene med ulike impulsresponslengder beregnes det 40 sett med L_2 -verdier og L_∞ -verdier. Disse plottes fra -90 til 90 grader, ved ingen elevasjon. Et gjennomsnitt av de 40 settene er beregnet og plottet med tykk rød strek.

I figur 4.10(a) vises de 40 settene med L_2 -verdier fra modeller med impulsresponser av lengde 10. Dette plottet er nærmest kaotisk, enhver realiser-

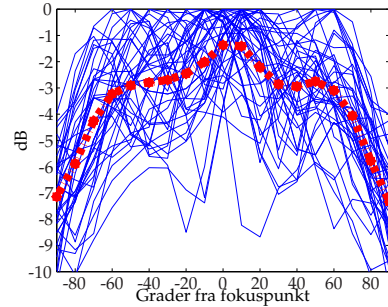


Figur 4.8: Fordeling av relative maksimale utslag for forskjellige impulsresponslengder. Den øverste fordelingen er fra modeller med impulsresponslengde 10, den nest øverste fra modeller med impulsresponslengde 100, den tredje er fra modeller med impulsresponslengde 200 og den siste fra modeller med impulsresponslengde 500.

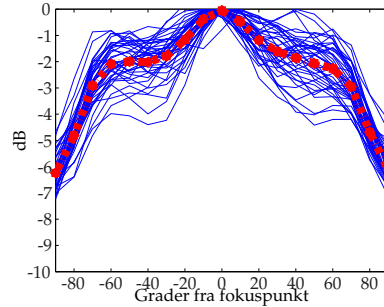


Figur 4.9: Fordeling av tilpasningsfeil ved forskjellige impulsresponslengder. Den øverste fordelingen er fra modeller med impulsresponslengde 10, den nest øverste fra modeller med impulsresponslengde 100, den tredje er fra modeller med impulsresponslengde 200 og den siste fra modeller med impulsresponslengde 500.

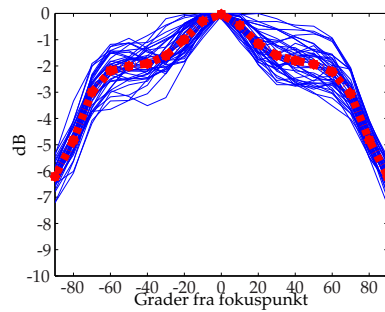
ing fraviker betraktelig fra gjennomsnittet. Det finnes realiseringer der det slippes igjennom mer energi i andre punkter enn fokuspunktet, noe som er uønsket. Etterhvert som impulsresponslengdene øker ser vi at for enhver modellrealisering, er ethvert sett med L_2 -verdier nærmere gjennomsnittet av settene. I figur 4.11 er de 40 settene med L_∞ -verdier plottet, fra modellene med ulike lengder på impulsresponsene. Det blir bedre fokusering med økte impulsresponslengder, i tillegg til at lengre impulsresponser bidrar til



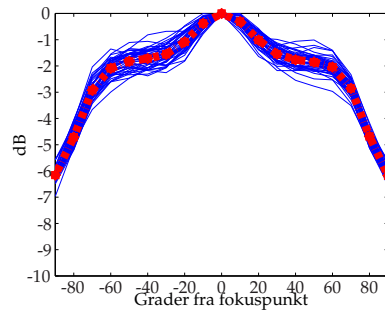
(a) L_2 -verdier plottet fra -90 til 90 grader fra fokuspunktet, ved ingen elevasjon, impulsresponslengde 10



(b) L_2 -verdier plottet fra -90 til 90 grader fra fokuspunktet, ved ingen elevasjon, impulsresponslengde 100



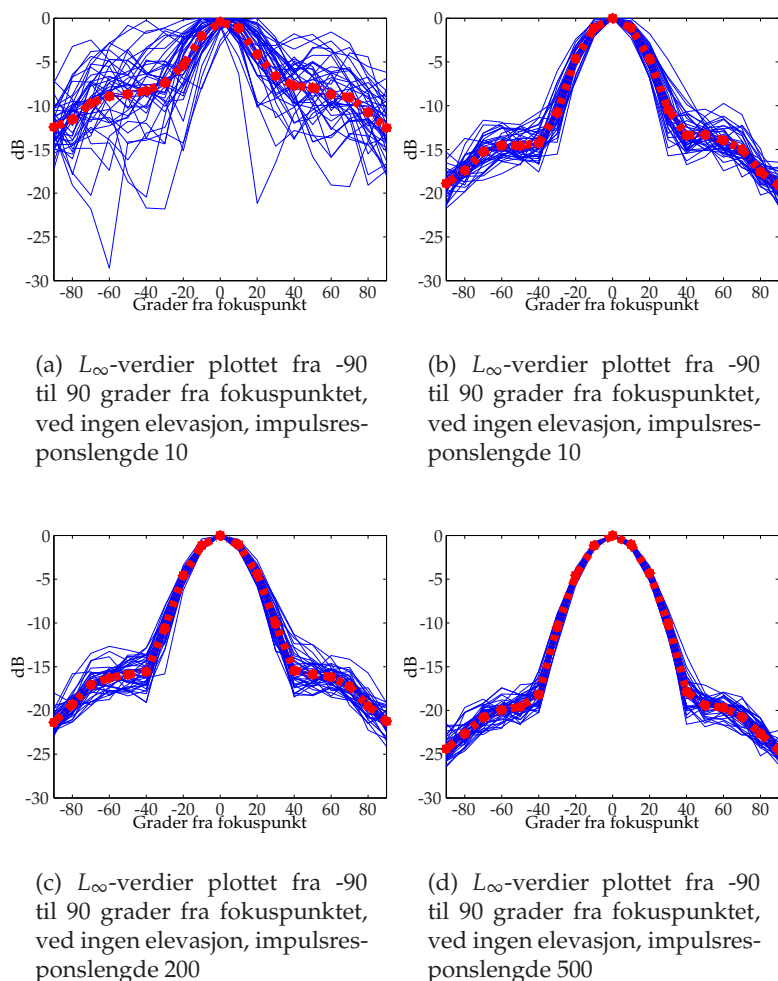
(c) L_2 -verdier plottet fra -90 til 90 grader fra fokuspunktet, ved ingen elevasjon, impulsresponslengde 200



(d) L_2 -verdier plottet fra -90 til 90 grader fra fokuspunktet, ved ingen elevasjon, impulsresponslengde 500

Figur 4.10: 40 plott av et utvalg av L_2 -verdier fra modeller med impulsresponser av lengde 10(a), 100(b), 200(c) og 500(d). De røde plottene er gjennomsnittene av de 40 realiseringene.

at ethvert sett med L_2 -verdier er nærmere gjennomsnittet. Sett i forhold til plottene i figur 4.10, ser vi en sammenheng mellom god fokusering med L_∞ -verdiene, og jevn fokusering med L_2 -verdiene. Avslutningsvis i dette delkapitlet betraktes enkelt-realiseringer av hele modellen med 2 forskjellige modell-oppsett:



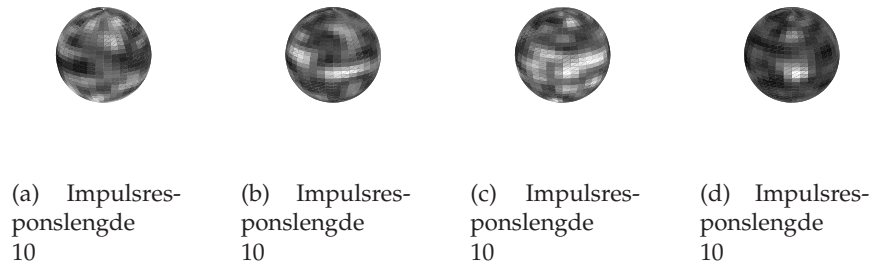
Figur 4.11: 40 plott av et utvalg av L_∞ -verdier fra modeller med impulsresponser av lengde 10(a), 100(b), 200(c) og 500(d). De røde plottene er gjennomsnittene av de 40 realiseringene.

- **Modell-oppsett 1:**
4 lydkilder, ett fokuspunkt og 4 undertrykkelsespunkter.
- **Modell-oppsett 2:**
10 lydkilder, ett fokuspunkt og 10 undertrykkelsespunkter.

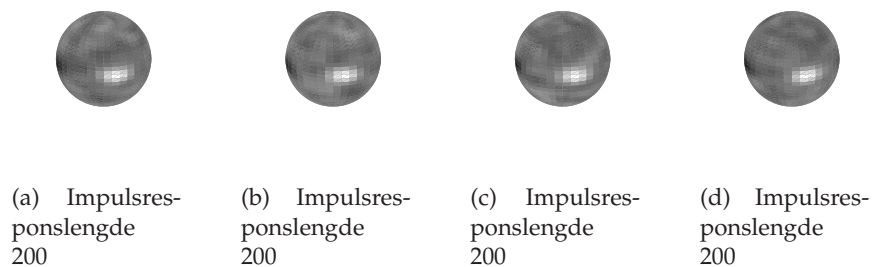
Av begge modelloppsettene realiseres det modeller med impulsresponser av lengde 10, og 200. I figur 4.12 og 4.13 er det plottet L_2 -verdier fra 4 enkelt realiseringer av modellen med modell-oppsett 1, fra modeller med henholdsvis impulsresponser av lengde 10 og 200. Plottene av L_2 -verdier

tilhørende modeller med impulsresponser av lengde 10 ser vi forandres fra realisering til realisering. I tre av plottene er det ikke noe synlig fokuspunkt. For modellrealiseringene med impulsresponser av lengde 200 ser vi at de fire realiseringene er likere hverandre. Plottene viser også distinkte fokuspunkt.

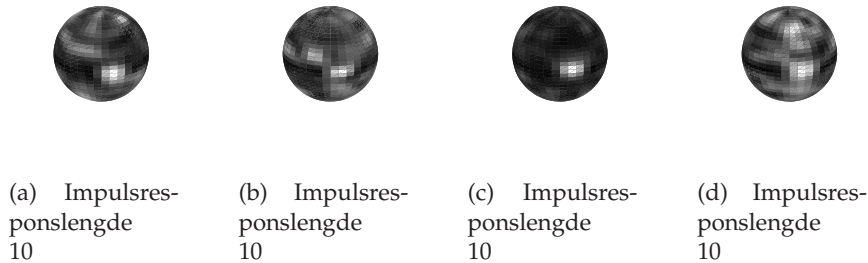
I figur 4.14 er L_2 -verdiene fra enkeltrealiseringer av modellen, med impulsresponser av lengde 10, med modell-oppsett 2 vist. Her ser vi også at plottene av L_2 -verdiene forandres fra realisering til realisering. Dersom vi øker lengden på impulsresponsene ser vi at L_2 -verdiene fra hver realisering blir likere hverandre, se figur 4.15.



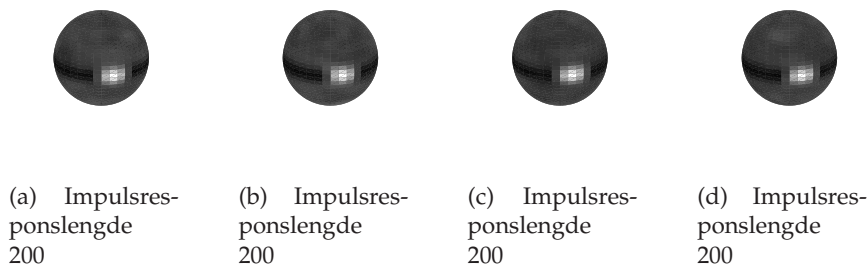
Figur 4.12: Plott av L_2 -verdiene ved enkelt realiseringer av modellen. Modellene har 4 lydkilder, ett fokuspunkt og 4 undertrykkelsespunkter og er påført glattingsskjema 1.



Figur 4.13: Plott av L_2 -verdiene ved enkelt realiseringer av modellen. Modellene har 4 lydkilder, ett fokuspunkt og 4 undertrykkelsespunkter og er påført glattingsskjema 1.



Figur 4.14: Plott av L_2 -verdiene ved enkelt realiseringer av modellen. Modellene har 10 lydkilder, ett fokuspunkt og 10 undertrykkelsespunkter og er påført glattingsskjema 1.



Figur 4.15: Plott av L_2 -verdiene ved enkelt realiseringer av modellen. Modellene har 10 lydkilder, ett fokuspunkt og 10 undertrykkelsespunkter og er påført glattingsskjema 1

Kommentar til dette delkapitlet

Ut i fra resultatene i dette delkapitlet kan det trekkes en empirisk slutning: Økte lengder på impulsreponsene gir bedre kontroll over lydfeltet som helhet, samt mer stabilitet i de relative energinivåene, de relative maksimale utslagene, mellom fokuspunktet og undertrykkelsespunktene, og tilpassningsfeilen ϵ . Vi har også sett en sammenheng mellom god fokusering med L_∞ -verdier og jevn fokusering med L_2 -verdier.

4.4 Avveining mellom faktisk fokusering og utilsiktet undertrykking

I det påfølgende vises det at man kan foreta en avveining mellom faktisk fokusering og utilsiktet undertrykking.

4.4.1 Forholdet mellom antall lydkilder og antall undertrykkelsespunkter og innvirkning på faktisk fokusering og utilsiktet undertrykking

Ved å gjøre om på forholdet mellom antall lydkilder og undertrykkelsespunkter kan man foreta en avveining mellom faktisk fokusering og utilsiktet undertrykking. Jeg har undersøkt forhold der antall lydkilder ikke overgår antall kontrollpunkter på kulen (fokuspunkt og undertrykkelsespunkter). Det viser seg at med færre undertrykkelsespunkter er den faktiske fokuseringen best, mens det da er mindre utilsiktet undertrykking. Dersom antallet undertrykkelsespunkter økes, forringes de relative energinivåene og tilpasningsfeilen noe, samtidig som den utilsiktede undertrykkingen bedres.

I de påfølgende simuleringene ser vi på et relativt energinivå, mellom fokuspunktet $(0,0)$ og ett undertrykkelsespunkt $(90,0)$, $f_2(d_1(n), d_2(n))$. $d_1(n)$ er responsen i fokuspunktet og $d_2(n)$ er responsen i undertrykkelsespunktet. Vi ser også på den utilsiktede undertrykking, f_2^- og tilpasningsfeilen ϵ , knyttet til det inverse problemet. Verdiene er gjennomsnittsverdier fra 100 modellrealiseringer. Modeller med 2 og 4 lydkilder betraktes. De gjennomsnittlige relative energinivåene, mellom fokuspunkt og undertrykkelsespunktet, fra modellene med 2 og 4 lydkilder vises henholdsvis i tabellene 4.4 og 4.5. I disse tabellene vises det at en avveiningen kan foretas mellom faktisk fokusering og utilsiktet undertrykking. De gjennomsnittlige L_2 -verdiene fra modellene med 2 og 4 lydkilder er plottet henholdsvis i figurene 4.16 og 4.17.

| # Lydkilder | # Undertrykkelsespunkter | $f_2(d_1(n), d_2(n))$ | f_2^- | ϵ |
|-------------|--------------------------|-----------------------|---------|------------|
| 2 | 1 | 8.3dB | 0.1dB | -5.7dB |
| 2 | 3 | 5.8dB | 2.6dB | -2.4dB |
| 2 | 4 | 5.5dB | 3.1dB | -1.7dB |
| 2 | 5 | 5.3dB | 3.3dB | -1.5dB |

Tabell 4.4: Gjennomsnittlige relative energinivåer, utilsiktet undertrykking og tilpasningsfeil for modeller med 2 lydkilder og med forskjellige antall undertrykkelsespunkter

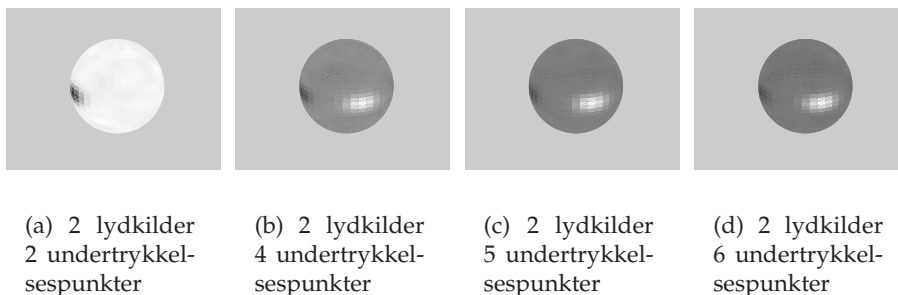
4.4 AVVEINING MELLOM FAKTISK FOKUSERING OG UTILSIKTET UNDERTRYKKING

| # Lydkilder | # Undertrykkelsespunkter | $f_2(d_1(n), d_2(n))$ | f_2^- | ϵ |
|-------------|--------------------------|-----------------------|---------|------------|
| 4 | 3 | 10.6dB | 2.6dB | -5.8dB |
| 4 | 5 | 9.2dB | 3.9dB | -3.6dB |
| 4 | 6 | 8.7dB | 4.3dB | -2.9dB |
| 4 | 7 | 8.4dB | 4.8dB | -2.4dB |

Tabell 4.5: Gjennomsnittlige relative energinivåer, utilsiktet undertrykking og tilpasningsfeil for modeller med 4 lydkilder med forskjellige antall undertrykkelsespunkter

Disse simuleringene viser en tilsynelatende god stabilitet i modellen, med tanke på antall undertrykkelsespunkter i forhold til den faktiske fokuseringen. Et problem er at tilpasningsfeilen ϵ øker for hvert undertrykkelsespunkt som legges til. Dette gjelder både for modellen med 2 og 4 lydkilder. Det oppnås en bedring i utilsiktet undertrykking, men et tap i både tilpasningsfeil og faktisk fokusering.

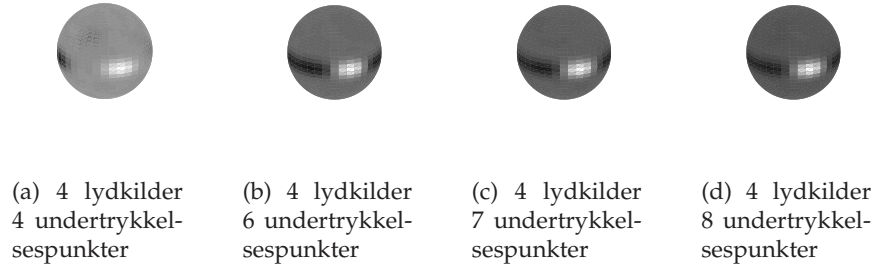
Dette kan tolkes som at antall frihetsgrader i systemet “brukes” opp, og man må inngå kompromisser for å klare å nå flere mål samtidig.



Figur 4.16: Gjennomsnittlige L_2 -verdier plottet fra modeller med 2 lydkilder og forskjellige antall undertrykkelsespunkter.

4.4.2 Lengden på de inverse filterne og påvirkning på de globale fokuseringsegenskapene til modellen.

Det er kjent at om man skal lage inverse filtere til en prosess, kan man oppnå bedre resultater om man øker lengden på det inverse filteret[15]. Å øke lengden på de inverse filterne viser seg også å påvirke forholdet mellom faktisk fokusering og utilsiktet undertrykking.



Figur 4.17: Gjennomsnittlige L_2 -verdier plottet fra modeller med 4 lydkilder og forskjellige antall undertrykkel-sespunkter.

| # Lydkilder | # Undertrykkel-sespunkter | Filterlengde | $f_2(.,.)$ | f_2^- | ϵ |
|-------------|---------------------------|--------------|------------|---------|------------|
| 2 | 2 | 2 | 7.2dB | 0.5dB | -4.2dB |
| 2 | 2 | 3 | 7.6dB | -0.1dB | -4.6dB |
| 2 | 2 | 4 | 7.7dB | -0.5dB | -4.7dB |
| 4 | 4 | 2 | 11.5dB | 2.1dB | -5.7dB |
| 4 | 4 | 3 | 12.5dB | 1.1dB | -6.4dB |
| 4 | 4 | 4 | 12.8dB | 0.5dB | -6.7dB |

Tabell 4.6: Gjennomsnittlige relative energinivåer, relative maksimale utslag og tilpasningsfeil for simuleringer hvor de inverse filterene er lengre enn impulsresponsene.

I tabell 4.6 er det relative energinivået mellom fokuspunktet (0,0) og ett undertrykkel-sespunkt, plassert i punktet (90,0) vist. Tabellen inneholder også den utilsiktede undertrykkingen og tilpasningsfeilen ϵ , som også her en gjennomsnittsverdier fra 100 modell-realiseringer. Jeg har sett på lengder av de inverse filterne som er 2, 3, og 4 ganger så lange som impulsresponsene. Denne tabellen viser også en avveining mellom tilpasningsfeil og faktisk fokusering, og utilsiktet undertrykking. Den utilsiktede undertrykkingen minker etterhvert som den faktiske fokuseringen og tilpasningsfeilen bedres.

4.5 Plasseringen av undertrykkel-sespunktene og innvirkning på utilsiktet undertrykking

I seksjon 4.2 så vi hvordan impulsreponser med lik tidsdomenerepresentasjon fra lydkildene til kontrollpunktene på kulen (fokuspunkt og undertrykkel-sespunkter) virket inn på de faktiske fokuseringsegneska-

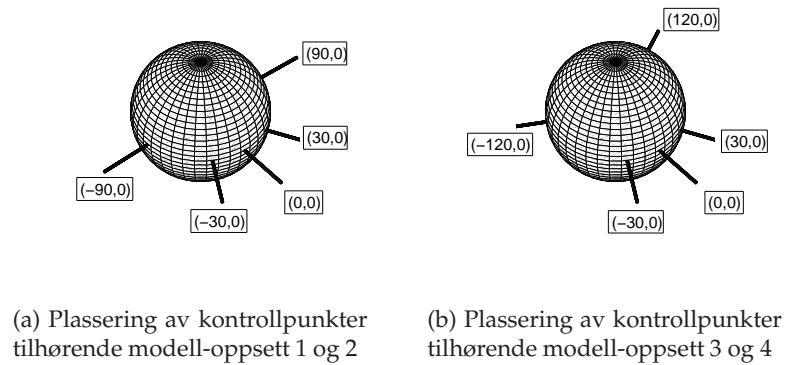
pene til modellen. Vi skal nå se at undertrykkelsepunktene plassering og mengden glatting også virker inn på de globale fokuseringsegenskapene til modellen.

I dette delkapitlet betrakter vi først modeller, med to forskjellige oppsett av undertrykkelsepunkter, og to forskjellige mengder glatting, det vil si fire forskjellige modell-oppsett.

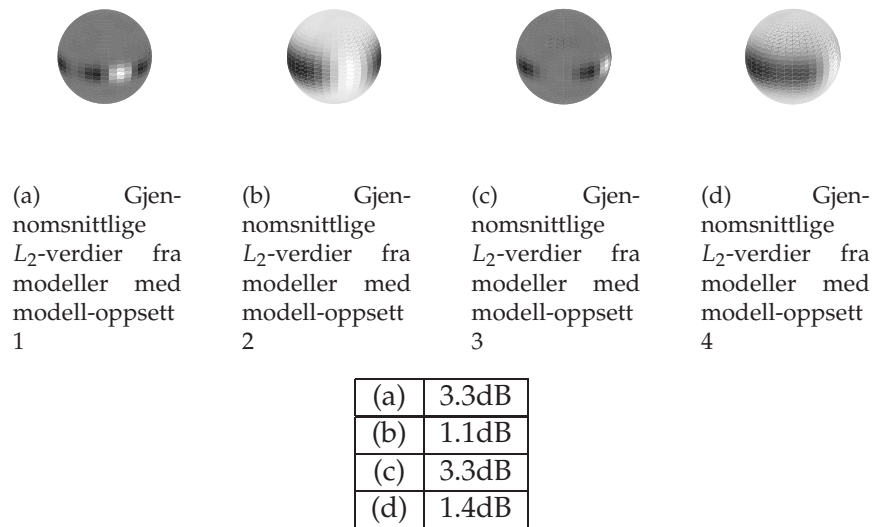
- **Modell-oppsett 1:**
4 lydkilder, 4 undertrykkelsepunkter og ett fokuspunkt. Fokuspunktet er plassert i punktet $(0,0)$, og undertrykkelsepunktene er plassert i $(90,0)$, $(-90,0)$, $(30,0)$ og $(-30,0)$. Disse modellene glattes med mønster 1, 3 ganger.
- **Modell-oppsett 2:**
4 lydkilder, 4 undertrykkelsepunkter og ett fokuspunkt. Fokuspunktet er plassert i punktet $(0,0)$, og undertrykkelsepunktene er plassert i $(90,0)$, $(-90,0)$, $(30,0)$ og $(-30,0)$. Disse modellene glattes med mønster 2, 6 ganger.
- **Modell-oppsett 3:**
4 lydkilder, 4 undertrykkelsepunkter og ett fokuspunkt. Fokuspunktet er plassert i punktet $(0,0)$, og undertrykkelsepunktene er plassert i $(120,0)$, $(-120,0)$, $(30,0)$ og $(-30,0)$. Disse modellene glattes med mønster 1, 3 ganger.
- **Modell-oppsett 4:**
4 lydkilder, 4 undertrykkelsepunkter og ett fokuspunkt. Fokuspunktet er plassert i punktet $(0,0)$, og undertrykkelsepunktene er plassert i $(120,0)$, $(-120,0)$, $(30,0)$ og $(-30,0)$. Disse modellene glattes med mønster 2, 6 ganger.

Plasseringen av kontrollpunktene på kulen kan ses i figur 4.18. For de fire forskjellige modell-oppsettene beregnes det gjennomsnittlige sett med L_2 verdier, av 100 modell-realiserings. De gjennomsnittlige L_2 -verdiene er plottet i figur 4.19

I plottene som er fremkommet fra modell-oppsettene med mest glatting er to av undertrykkelsepunktene plassert slik at de påvirkes av glattingen nær fokuspunktet, jamfør delkapittel 4.2. Som vi kan se er det mindre distinkte fokuspunkt i disse to plottene, se figurene 4.19(b) og 4.19(d). I figurene som har fremkommet fra modellrealiseringer med mindre glatting ser vi distinkte fokuspunkt. I den tilhørende tabellen i figur 4.19 ser vi også mye av den utilsiktede undertrykkingen går tapt ved modell-oppsettene med mest glatting.



Figur 4.18: Plassering av kontrollpunktene i modell-oppsettene 1,2,3 og 4.



Figur 4.19: Plott av gjennomsnittlige L_2 -verdier fra modeller med modell-oppsett 1, 2, 3, og 4, med tilhørende gjennomsnittlige verdier på utilsiktet undertrykking.

Denne mangelen på fokuspunkt, i modellene med mest glatting, forekommer ikke bare når halvparten av undertrykkelsespunktene plasseres for nære fokuspunktet. Vi ser derfor også på en modell med seks lydkilder og seks undertrykkelsespunkter. To av undertrykkelsespunktene plasseres slik at de påvirkes av glattingen nær fokuspunktet. Ytterligere to modell-

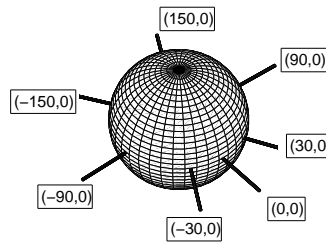
oppsett betraktes.

- **Modell-oppsett 5:**

6 lydkilder, 6 undertrykkelsespunkter og ett fokuspunkt. Fokuspunktet er plassert i punktet $(0,0)$, og undertrykkelsespunktene er plassert i punktene $(30,0)$ $(-30,0)$, $(90,0)$, $(-90,0)$ $(150,0)$ og $(-150,0)$. Disse modellene påføres glattingsmønster 1, 3 ganger.

- **Modell-oppsett 6:**

6 lydkilder, 6 undertrykkelsespunkter og ett fokuspunkt. Fokuspunktet er plassert i punktet $(0,0)$, og undertrykkelsespunktene er plassert i punktene $(30,0)$ $(-30,0)$, $(90,0)$, $(-90,0)$ $(150,0)$ og $(-150,0)$. Disse modellene påføres glattingsmønster 2, 6 ganger.



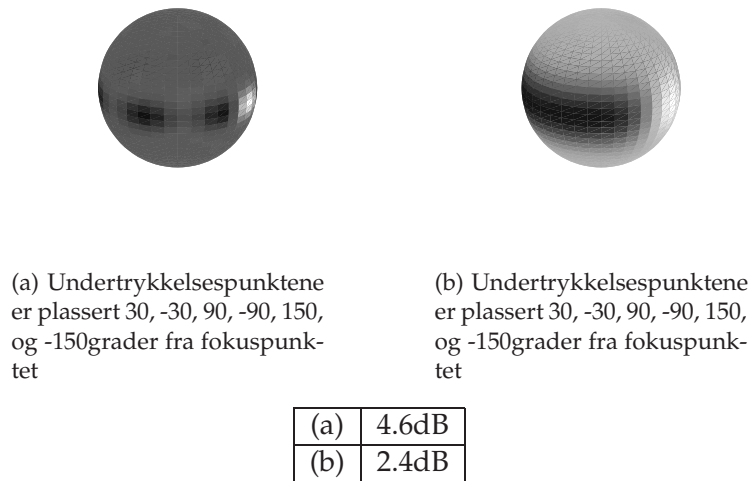
Figur 4.20: Plassering av kontrollpunktene i modell-oppsettene 5 og 6.

Plasseringen av kontrollpunktene på kulen kan ses i figur 4.20. Gjennomsnittlige L_2 -verdier fra 100 modell-realiseringer av modell-oppsettene 5 og 6 er plottet i 4.21. Disse plottene viser de samme tendensene som plottene i de forrige simuleringene, de globale fokuseringsegenskapene til modellen forringes om undertrykkelsespunkter plasseres for nærme fokuspunktet. Verdiene for den utilsiktede undertrykkingen er vist i tilhørende tabell i figur 4.21.

Det bør påpekes at den utilsiktede undertrykkingen også ville forringes i modellen med mindre glatting, undertrykkelsespunktene måtte bare ha vært plassert nærmere fokuspunktet. Dette kan ses på som en slags robusthet ved modeller som er påført mindre glatting, undertrykkelsespunktene kan plasseres mer fritt på kulen uten at det påvirker de globale fokuseringsegenskapene til modellen.

4.6 Oppsummering

Forsøkene utført i dette kapitlet viser at innholdsrike impulsresponser øker fokuseringsegenskapene til modellen. Vi har sett at hvor godt et lokalt lyd-



Figur 4.21: Plott av gjennomsnittlige L_2 -verdier fra modeller med modelloppsett 5 og 6, med tilhørende gjennomsnittlige utilsiktede undertrykkingsverdier.

felt kan kontrolleres avhenger av rikdommen til impulsresponsene. Det er også en korrespondanse mellom *jevn fokusering* med L_2 -verdiene og *god fokusering* med L_∞ -verdiene. Videre har vi sett at det er en avveining mellom faktisk fokusering og utilsiktet undertrykking. Simuleringene viser at impulsresponser med innholdsrik tidsdomene-representasjon fra lydkildene til undertrykkelsespunktene, er viktig både for den faktiske fokuseringen, størrelsen på fokuspunktet og den utilsiktede undertrykkingen. Med tanke på de relative energinivåene vil disse bare kunne gjøres bedre med en økning av antall lydkilder. Dette er viktig om systemet skal benyttes til applikasjoner for hørbar lyd. Det kan være det finnes applikasjoner der man kan utnytte det faktum at de relative maksimale utslagene bedres med mer innholdsrike impulsresponser uten en økning av antall lydkilder.

Kapittel 5

Eksperimenter

Det forrige kapitlet viste interessante resultater. Disse følges opp med eksperimenter på ekte data. De ekte dataene består av impulsresponser målt i et rom. Jeg har forsøkt å anskaffe innholdsrike impulsresponser ved å sette reflektorer foran høyttalerne. Et sett med impulsresponser målt uten reflektorene foran høyttalerne benyttes også. Lydfokusering med begge disse settene med impulsresponser sammenliknes. Resultatene av denne simuleringen stemmer overens med resultater fra forrige kapittel. Fokuseringen med impulsresponsene målt med reflektorer gir tilsynelatende mer kontroll over lydfeltet som helhet. Robustheten til systemet, hvor det legges til støy på impulsresponsene, undersøkes også.

Slutten av dette kapitlet tar for seg kildeseparasjon og lydfokusering med hode-relaterte impulsresponser. Kildene plasseres virtuelt rundt hode til en person ved at kildene filtreres av hode-relaterte impulsresponser tilhørende disse posisjonene. Rammeverket for kildeseparasjon, gjennomgått i kapittel 2, brukes for å sette opp et likningsystem for dette formålet. Løsningen av dette likningsystemet er optimale filtre som kan brukes til å filtrere "utgangene" av hvert øret. Ved lydfokusering med hode-relaterte impulsresponser kan vi tenke oss at test personen har høyttalere inne i øret.

5.1 Undersøkelse av lydfokusering med 2 høyttalere i rom

I dette delkapitlet undersøkes lydfokusering med impulsresponser målt i et rom. To sett med impulsresponser benyttes. Begge settene består av impulsresponser målt fra to høyttalere i 15 punkter. Ved målingen av det ene settet har jeg satt reflektorer foran høyttalerne, i et forsøk på å oppnå mer innholdsrike impulsresponser. Reflektorene er tomme glassflasker, se



Figur 5.1: Bilde av oppsettet

figur 5.1. Dette er et forsøk på å oppnå mer innholdsrike impulsresponsen. Det andre settet med impulsresponsen er målt uten reflektorene.

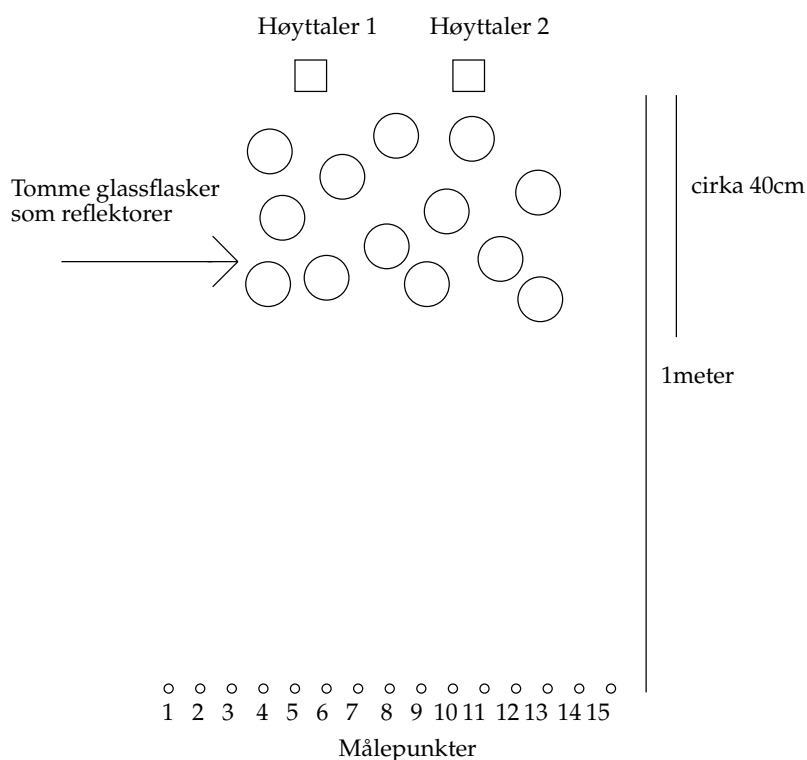
5.1.1 Oppsettet av systemet

Jeg har brukt 2 høyttalere og en mikrofon for å måle impulsresponsen. Impulsresponsene er målt mellom høyttaler 1 og 2, og 15 punkter. Dette resulterer i ett sett med 30 impulsresponsen, to impulsresponsen tilhørende hvert punkt. I figur 5.2 vises en illustrasjon av oppsettet.

5.1.2 Måling av impulsresponsen

For å måle impulsresponsene fra høyttalerne til punktene brukes et sinus-signal med eksponensielt økende frekvens. Detaljene om denne metoden for å måle impulsresponsen på finnes i tillegg A. Målingene gjøres med et program skrevet i Matlab™, med programpakken SND. Denne programpakken gjør det mulig å synkront spille av et målesignal og samtidig gjøre et opptak av signalet med en mikrofon. Målingene lagres på en datamaskin for å så å bli konvolvert med det inverse filteret til målesignalet. Målingene utføres med en samplerate på 44.1KHz.

Jeg er interessert i impulsresponsene til det "lokale området" foran høyttalerne, som om målingene var blitt gjort i et lyddempet rom. Rommet målingene ble utført i var ikke lyddempet, dermed må impulsresponsene tilhørende dette "lokale området" estimeres ved å skjære vekk deler av de målte



Figur 5.2: Oppsettet av systemet ved målinger av impulsresponsen i rom.

impulsresponsene.

Lydens hastighet i luft er cirka $344 \frac{\text{meter}}{\text{sekund}}$. Et sample med en samplerate på 44.1KHz tilsvarer cirka $0.0078 \frac{\text{meter}}{\text{sample}}$. 130 sampler tilsvarer tiden det tar for lyden å forflytte seg en meter, $130 \text{ sampler} \times 0.0078 \frac{\text{meter}}{\text{sample}} \approx 1\text{m}$. I tillegg til forsinkelsene mellom målepunktene og høyttalerne antar jeg at ytterligere 150 samples er tilstrekkelig for å inkludere det lokale området foran høyttalerne.

Antall sampler det trengs for å få med seg lyden som propagerer gjennom reflektorene er cirka 50 samples da reflektorene har en utstrekning på cirka 40cm, $50 \text{ sampler} \times 0.0078 \frac{\text{meter}}{\text{sample}} \approx 0.4\text{m}$. Jeg antar at tre ganger denne tiden er tilstrekkelig for å få med refleksjonene fra reflektorene, altså cirka 150 samples. Dette resulterer i impulsresponsen med tilsammen $130 + 150 = 280$ filtertapper.

Antall samples som inkluderes fra reflektorene vil variere litt i forhold fra punkt til punkt. Dette er fordi det vil være litt forskjellige strekninger mel-

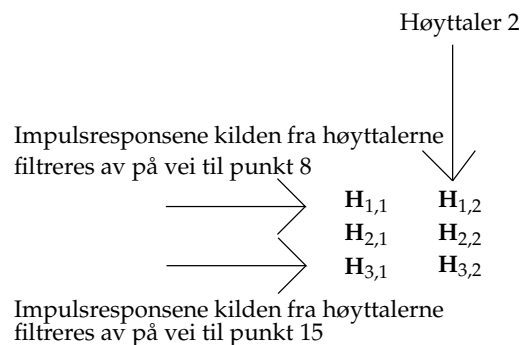
lom høyttalerne og målepunktene. Lydens reisevei er litt lenger fra høyttaler 1 til punkt 15, enn fra høyttaler 1 til punkt 5.

Begge settene med impulsresponsen inneholder antageligvis noen refleksjoner fra rommet også, siden rommet ikke var lyddempet.

5.1.3 Simuleringer med impulsresponsen målt i rom: Sammenlikning av fokusering med impulsresponsen målt med og uten reflektorer

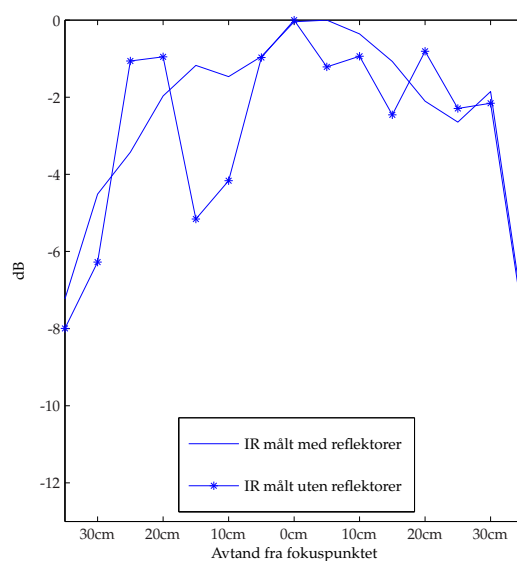
Jeg velger punkt 8 i figur 5.2 som fokuspunkt, og punkt 1 og 15 som undertrykkelsespunkter. La impulsresponsen fra høyttaler 1 til punkt 8, være $h_{1,1}(n)$, impulsresponsen fra høyttaler 2 til punkt 8 være $h_{1,2}(n)$. La impulsresponsene mellom høyttaler 1 og 2, og punkt 1 være henholdsvis $h_{2,1}(n)$ og $h_{2,2}(n)$, og impulsresponsene mellom høyttaler 1 og 2, og punkt 15 være henholdsvis $h_{3,1}(n)$ og $h_{3,2}(n)$

Konvolusjonsmatrisene tilhørende impulsresponsene målt i punkt 8 settes så inn i øverste blokk rad i den sammensatte konvolusjonsmatrisen \mathbf{H} , konvolusjonsmatrisene tilhørende impulsresponsene i punkt 1 settes inn i blokk-rad 2, og så videre, se figur 5.3. Deretter løses likningssystemet 2.21 for å konstruere de inverse-filterne $g_1(n)$ og $g_2(n)$. Filterne $g_1(n)$ og $g_2(n)$ brukes til å filtrere impulsresponsene tilhørende alle de 15 tilgjengelige punktene. For hvert punkt p_i beregnes L_2 - og L_∞ -verdien. $L_2(p_i)$ er plottet i figur 5.4, $L_\infty(p_i)$ er plottet i figur 5.5.



Figur 5.3: Hjelpesfigur til sammensetning av den sammensatte konvolusjonsmatrisen

Vi ser først på figur 5.4. Denne figuren viser fokuseringen som oppnås med begge settene med impulsresponsen. Vi kan se at fokuseringen hvor impulsresponsene med reflektorene er brukt gir en jevnere fokusering. Fokuseringen følger mer en kurve. Ved fokusering med impulsresponsen uten



Figur 5.4: Plott av fokuseringen med impulsresponser med og uten reflektorer, L_2 verdiene plottet

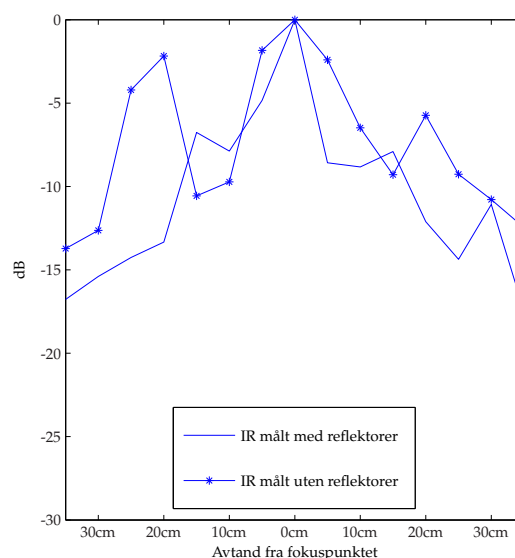
reflektorene ser vi at fokuseringen er mer ujevn. Det er vanskelig å se et mønster i dette plottet.

I figur 5.5 er L_∞ -verdiene plottet. Dette plottet viser en skarpere fokusering hvor impulsresponsene med målt med reflektorer er brukt. Det er også bedre undertrykking av undertrykkelsespunktene. Fokuseringen med impulsresponsene målt uten reflektorene gir også her en mer ujevn fokusering.

Begge disse plottene korresponderer med simuleringene i kapittel 4: Mer innholdsrike impulsresponser, gir jevnere fokusering med L_2 verdier, og bedre fokusering med L_∞ verdier. Figurene 5.4 og 5.5 kan sammenliknes med figurene 4.10 og 4.11 i kapittel 4, som viser en korrespondanse mellom de ekte og de simulerte impulsresponsene. Det ser ut til at bruk av reflektorer gir ønsket resultat, selv med relativt få reflektorer.

5.1.4 Robusthet/Støy-analyse

I det påfølgende skal undersøkes det hvor robust løsningen av likningsystemet 2.21 er mot støy. Dette er interessant siden de estimerte impulsresponsene vil inneholde en eller flere former for støy. Støyen kan komme fra måleinstrumentene (høytalere, mikrofon eller lydkort i PC), feil som oppstår ved diskretisering av dataene under sampling av målesignalene, støy i



Figur 5.5: Plott av fokuseringen med impulsresponser med og uten reflektorer, L_∞ verdiene plottet

rommet hvor målingene utføres og så videre. I det påfølgende representerer støyen fraviket fra de eksakte impulsresponsene.

Man benytter konvolusjonsmatrisene tilhørende impulsresponsestimater i matriselikningen 2.21 til å konstruere inverse filtere $g_1(n)$ og $g_2(n)$. Ved prefiltrering av høyttalerne med disse filterne fokuseres lyden i et punkt, samtidig som den skal undertrykkes i andre punkter. Om impulsresponsestimatene fraviker for mye fra de eksakte impulsresponsene, vil ikke de prefiltrerte utgangene lenger gi den ønskede romlige responsen. La de målte impulsresponsene som inngår i matriselikningen 2.21 være

$$\begin{aligned} h_{1,1}^{\text{målt}}(n) &= h_{1,1}(n) + \eta_1(n) \\ h_{1,2}^{\text{målt}}(n) &= h_{1,2}(n) + \eta_2(n) \\ &\vdots \\ h_{3,2}^{\text{målt}}(n) &= h_{3,2}(n) + \eta_6(n). \end{aligned}$$

Hvor $\eta_1(n), \eta_2(n), \dots, \eta_6(n)$ er støy. Konvolusjonsmatrisene tilhørende disse impulsresponsestimatene utgjør den sammensatte konvolusjonsmatrisen \mathbf{H} i 2.21 og kan skrives som

$$\mathbf{H}_{\text{målt}} = (\mathbf{H} + \mathbf{E}) \quad (5.1)$$

Effekten av dette vil være at løsningsfilterne, $g_1(n)$ og $g_2(n)$, vil være konstruert til å gi følgende respons

$$\begin{aligned} h_{1,1}^{\text{målt}}(n) * g_1(n) + h_{1,2}^{\text{målt}}(n) * g_2(n) &\approx \delta(n) \\ h_{2,1}^{\text{målt}}(n) * g_1(n) + h_{2,2}^{\text{målt}}(n) * g_2(n) &\approx 0(n) \\ h_{3,1}^{\text{målt}}(n) * g_1(n) + h_{3,2}^{\text{målt}}(n) * g_2(n) &\approx 0(n) \end{aligned}$$

Mens de burde vært konstruert til å gi

$$\begin{aligned} h_{1,1}(n) * g_1(n) + h_{1,2}(n) * g_2(n) &\approx \delta(n) \\ h_{2,1}(n) * g_1(n) + h_{2,2}(n) * g_2(n) &\approx 0(n) \\ h_{3,1}(n) * g_1(n) + h_{3,2}(n) * g_2(n) &\approx 0(n) \end{aligned}$$

Løsningsfilterne $g_1(n)$ og $g_2(n)$ vil dermed være konstruert til å gi en ønsket respons med impulsresponsen som fraviker fra de eksakte. Dersom de inverse filterne $g_1(n)$ og $g_2(n)$ benyttes til å prefiltrere utgangene av lydkildene og det er støy i omgivelsene der fokuseringen skal utføres, kan fraviket øke ytterligere.

I det påfølgende skal vi se hvor stort utslag støy på impulsresponsene vil gi med tanke på tilpasningsfeil, relative energinivåer og romlig respons.

5.1.5 Signal-støy-forhold

La $y(n)$ være et estimat av signal $x(n)$ med additiv støy $\eta(n)$. Da er signal-støy-forhold (SNR) definert som

$$\text{SNR} = 10 \log_{10} \frac{\|x(n)\|_2^2}{\|\eta(n)\|_2^2}, \quad (5.2)$$

her angitt i desibel. I situasjonen betraktet her har vi ikke ett signal med støy, vi har til hvert målepunkt relatert to impulsresponsen med støy. Konvolusjonsmatriser tilhørende 6 impulsresponsen utgjør den sammensatte konvolusjonsmatrisen \mathbf{H} i 2.21. Denne matrisen inngår i likningsystemet som løses for å estimere de optimale filterne. Det kan angis et signal-støy forhold, $\text{SNR}_1, \dots, \text{SNR}_6$ for hver impulsrespons. Jeg bruker derfor et SNR-mål som er et gjennomsnitt av de seks signal-støy forholdene.

$$\text{SNR}_{\text{system}} = \frac{\text{SNR}_1 + \dots + \text{SNR}_6}{6}. \quad (5.3)$$

Alle impulsresponsene har cirka like mye energi slik at hvert enkelt signal-støy forhold $\text{SNR}_1, \dots, \text{SNR}_6$ varierer i intervallet $[-1\text{dB}, 1\text{dB}]$ fra gjennomsnittsverdien 5.3. De forskjellige signal-støy forholdene jeg har betraktet er 0, 5, 10, 20, 30 og 40dB.

5.1.6 Tilpasningsfeil og relative energinivåer ved løsningen av likningssystemet med støy.

Punkt 8 i figur 5.2 velges igjen som fokuspunkt og punkt 1 og 15 som undertrykkelsespunkter. Den sammensatte konvolusjonsmatrisen \mathbf{H} settes sammen på samme måte som i seksjon 5.1.3. \mathbf{H} settes så inn i likningssystemet 2.21, og dette løses. For de valgte signal-støy forholdene undersøkes det hvordan løsningsfilterne $g_1(n)$ og $g_2(n)$ presterer når det legges støy på impulsresponsene.

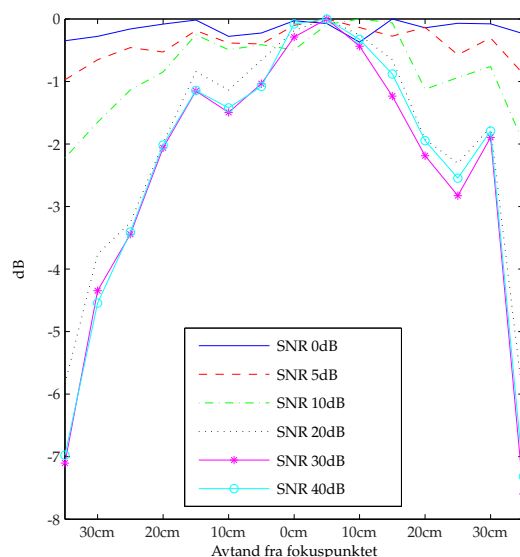
I Andre kolonne i tabell 5.1 vises tilpasningsfeilen ϵ ved de ulike signal-støy forholdene. Tilpasningsfeilen uten støy er -3.7dB. Ved lave SNR (0 - 10dB) kan vi se at løsningen gir veldig høy tilpasningsfeil i forhold til tilpasningsfeilen uten støy. Ved høyere signal-støy forhold (20 - 40dB) presterer løsningen mer tilfredstillende, med tanke på tilpasningsfeilen. Dette gjenspeiles også i tredje kolonne i tabell 5.1 som viser det relative energinivået mellom fokuspunktet og ett av undertrykkelsespunktene (punkt 15 i figur 5.2).

I figur 5.6 vises fokuseringen med L_2 -verdiene hvor forskjellig grad av støy er lagt til impulsresponsene. Ved lave signal-støy forhold har vi mistet all fokusering, mens ved bedre signal-støy forhold ser vi at fokuseringen er mer tilfredstillende (i forhold til fokusering uten støy på impulsresponsene).

De laveste signal-støy forholdene sett på her kan være realistiske i reelle situasjoner. Dette kan være situasjoner der man i tillegg til en viss usikkerhet i de målte impulsresponsene også må utføre fokuseringen i støyende omgivelser. Det finnes varianter av minste kvadraters metode som tar høyde for datamatriser konstruert fra usikre data. Dette er metoder som "Data Least Squares" [7] og "Total Least Squares" [18]. Båndstrukturen på matrisene jeg har jobbet med i denne oppgaven gjør imidlertid disse metodene relativt kompliserte, disse metodene må derfor forfølges i et senere arbeid.

| SNR | ϵ | $f_2(.,.)$ |
|------|------------|------------|
| 0dB | 12.3dB | 0.2dB |
| 5dB | 7.5dB | 0.8dB |
| 10dB | 3.2dB | 1.4dB |
| 20dB | -2.2dB | 5.9dB |
| 30dB | -3.5dB | 7.3dB |
| 40dB | -3.5dB | 7.2dB |

Tabell 5.1: Tilpasningsfeilen og relative energinivåer ved fokusering under støyete forhold



Figur 5.6: L_2 -verdiene plottet hvor det er lagt støy på impulsresponsene

5.2 Kildeseparasjon: Undersøkelse av separasjonsegenskaper med hode-relaterete-impulsresponser

I dette delkapitlet utdypes resultatene fra artikkelen “Source Enhancement and Suppression with Binaural Head Related Impulse Response Functions”, vedlagt i tillegg B, ved at de settes i sammenheng med resultater fra forrige kapittel.

5.2.1 Et sett med hode-relaterte impulsresponser

I de påfølgende simuleringene benyttes en database med hode-relaterte impulsresponser (HRIR). Databasen inneholder hode-relaterte impulsresponser tilhørende 7 testpersoner, og dokumenteres i [12]. I disse simuleringene har jeg benyttet meg av hode-relaterte impulsresponser tilhørende en testperson. Punktene impulsresponsene måles fra er plassert på et kule-aktig gitter rundt lytteren. Til hvert av de 1093 punktene er det relatert to impulsresponser. En impulsrespons fra punktet til hver av ørene.

Impulsresponsene er gyldige i frekvensintervallet 0 til 21kHz. De er samlet med en samplerate på 83.3kHz for etterpå å bli lavpassfiltrert med 4. ordens Bessel-filter med kutt-frekvens på 18kHz. På grunn av begrensinger i måleinstrumentene er HRIRene blitt tilpasset i de lave (0-700Hz), og høye (18-21kHz) frekvensene. De hode-relaterte impulsresponsene er relatert til posisjoner om testpersonen, gitt av polar-koordinater, ϕ angir posisjonen i det horisontale planet, θ i det vertikale planet. $(\phi, \theta) = (0, 0)$ korresponderer til en kilde rett foran lytteren, $(\phi, \theta) = (0, 90)$ korresponderer til en kilde rett over lytteren.

5.2.2 Simuleringer med hode-relaterte impulsresponser

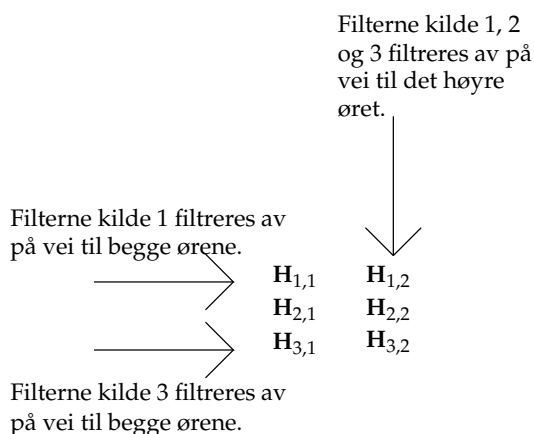
Kildeplassering 1

Anta at vi har en ønsket kilde i punktet (85,0), kall den kilde A. Anta at vi har to uønskede kilder, i punktene (0,0) og (-85,0), kall disse henholdsvis kilde B og C. Kilden plassert i punktet (85,0) vil filtreres av impulsresponsene tilhørende dette punktet før lyden når trommehinnene i ørene. Kilden plassert i punktet (0,0) filtreres av impulsresponsene tilhørende dette punktet og så videre. Målet er å kunne separere ut den ønskede kilden, kilde 1, fra de to uønskede.

La $\mathbf{H}_{1,1}$ være konvolusjonsmatrisen tilhørende impulsresponsen fra punktet (85,0) til det venstre øret og la $\mathbf{H}_{1,2}$ være konvolusjonsmatrisen tilhørende impulsresponsen fra punktet (85,0) til det høyre øret. Disse to matrisene plasseres i øverste blokk-rad i den sammensatte konvolusjonsmatrisen \mathbf{H} . Konvolusjonsmatrisene tilhørende impulsresponsene til de to uønskede kildene utgjør blokk-rad 2 og 3, se figur 5.7.

Sett \mathbf{H} inn i likningsystemet 2.21 og løs dette. Løsningsvektoren \mathbf{g} vil inneholde de optimale separasjons-filterne $g_1(n)$ og $g_2(n)$. Dersom $g_1(n)$ og $g_2(n)$ brukes til å filtrere "utgangene" av henholdsvis venstre og høyre øre, vil kilde 1, ideelt sett separeres fra kilde 2 og 3.

I forrige kapittel ble fokuseringen visualisert ved å filtrere impulsresponsene tilhørende alle tilgjengelige punkter med de konstruerte inverse filterne



Figur 5.7: Hjelpesfigurer til sammensetningen av matrisen \mathbf{H} ved kilde-separasjon med HRIR.

$g_1(n)$ og $g_2(n)$ (hvor de inverse filterne var konstruert for å gi en ønsket respons i noen utvalgte punkter). For å visualisere separasjonen benyttes en helt tilsvarende framgangsmåte her, for ethvert tilgjengelig punkt beregner $L_2(\phi_i, \theta_i)$, ved å filtrere impulsresponsene tilhørende hvert punkt med løsningsfilterne $g_1(n)$ og $g_2(n)$.

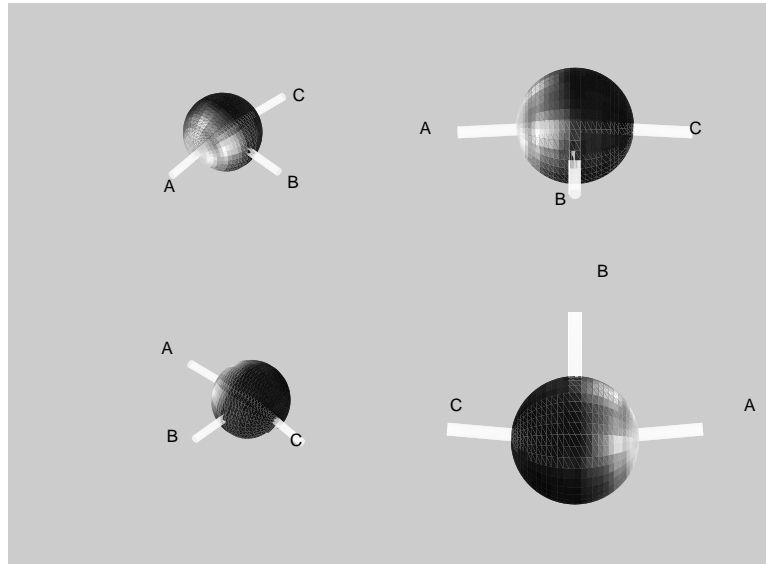
Punktene rundt testpersonen impulsresponsene er målt fra, er ujevnt plassert. For å lettere å visualisere resultatet er det benyttet en bi-lineær interpolasjon over settet $L_2(\phi_i, \theta_i)$ for å estimere verdier i punkter der det ikke finnes noen verdier.

Kuleplott med hode-relaterte impulsrespons

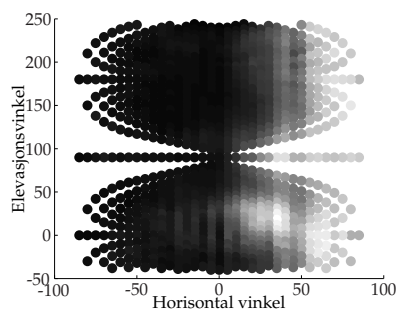
Resultatet av prosessen beskrevet ovenfor kan sees i figur 5.8. Den samme kule vises fra 4 forskjellige vinkler. Sylindren merket A skjærer kule der den ønskede kilden er, sylindrene merket B og C skjærer kule der de uønskede kildene er. Vi ser en tilsynelatende god fokusering inn på den ønskede kilden, samtidig som de uønskede undertrykkes. Denne kildeoppstillingen gir $\text{SIR}_{\text{ut}} = 7.8\text{dB}$.

Et slikt plott i kontekst av kilde-separasjon kan ikke tolkes som at alle punkter på kule er potensielle lydkilder. Dette ville forringet separasjonsverdien SIR_{ut} fullstendig. Plottet kan imidlertid gi et inntrykk av hvor presis kilde-separasjonen forblir ved bevegelse av kildene. Betrakt figur 5.8, den ønskede kilden A kunne ha blitt flyttet rundt på et relativt stort område uten at separasjonen ville blitt forringet. Det samme gjelder kilde C. Kilde B derimot kan flyttes litt mot kilde A og den vil straks gjøre et større utslag,

og SIR_{ut} vil da forandres. Derimot om den forflyttet seg mot kilde C ville separasjonsverdien SIR_{ut} tilsynelatende ikke forandres nevneverdig.



Figur 5.8: "Hode-plott" med fokus inn på en kilde i punktet (85,0). Sylinderen merket A, skjærer kulen der den ønskede kilden er plassert. Sylinderne merket B og C skjærer kulen der de uønskede kildene er plassert.



Figur 5.9: Overblikk av separasjonen med hode-relaterte impulsresponser, ved kildeplassering 1.

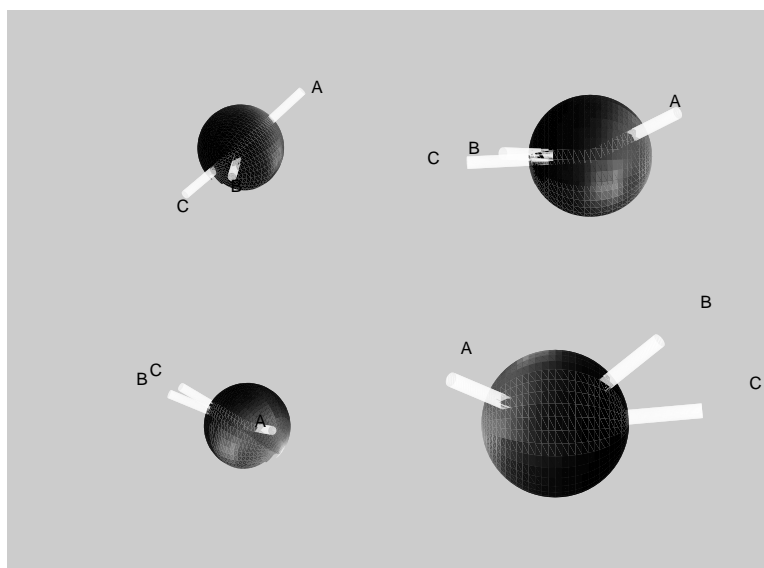
Kildeplassering 2

I neste simulering har vi følgende kildeplassering: Vi har en ønsket kilde i punktet (-45,65) og to uønskede kilder plassert i punktene (85,0) og (45,37).

Vi går fram på samme måte som ved forrige simulering og løser liknings-systemet 2.21. Dette gir oss løsningsfilterne $g_1(n)$ og $g_2(n)$ som brukes til å anskaffe settet $L_2(\phi_i, \theta_i)$. Visualiseringsprosessen er tilsvarende som ved forrige simulering.

Denne oppstillingen gir $SIR_{ut} = 14.9\text{dB}$, som er nærmest en fordobling av verdien ved kildeplassering 1. Kan denne verdien skyldes at det i utgangspunktet slippes igjennom mer energi fra dette punktet?

Om ingen inversfiltrering gjøres (om de inverse filterne velges til å være diskrete delta-funksjoner av endelig lengde) vil det i denne kildeoppstillingen slippe gjennom 2.3dB fra fokuspunktet. Fra de to uønskede kildene vil det slippe gjennom 3dB og 2.7dB. De inverse filterne gjør altså en betydelig forskjell. Kuleplott tilhørende denne simuleringen kan sees i figur 5.10.



Figur 5.10: "Hodeplott" med fokus inn på en kilde plassert i punktet (-45,65). Sylindren merket A, skjærer kulen der den ønskede kilden er. Sylindrene merket B og C skjærer kulen der de uønskede kildene er.

Denne figuren viser også en relativt god robusthet mot bevegelse av den ønskede og de uønskede kildene. Den halvdelen av kulen hvor de uønskede kildene er plassert er mørk, slik at om de uønskede kildene forflyttes rundt på denne halvdelen vil de tilsynelatende fortsatt være godt undertrykt. Igjen er det slik at dersom de uønskede kildene beveges mot den ønskede kilden vil de straks gjøre ett større utslag og

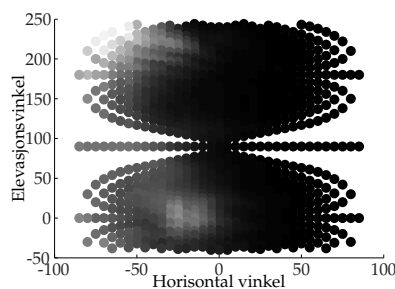
separasjonsverdien SIR_{ut} vil forringes.

Ustabil fokusering med hode-relaterte impulsresponser

Når man måler hode-relaterte impulsresponser plasserer man vanligvis en veldig liten mikrofon inne i øret til en person. Deretter spiller man av et passende målesignal med en høyttaler. Signalet tas opp av mikrofonen og kan ved riktig prosessering gjøres om til en impulsrespons. Impulsresponsene måles fra forskjellige posisjoner ved at man flytter høyttaleren. I [38] benyttes en metode der mikrofon og høyttaler har byttet plass. En veldig liten høyttaler plasseres inne i øret til en person og impulsresponser måles i punkter rundt hodet med mikrofoner. Det vises i [38] at disse to metodene for å måle impulsresponser på gir cirka like resultater.

Det kan her da trekkes en parallell mellom kilde-separasjon og lydfokusering om vi tenker oss at personen de hode-relaterte impulsresponsene tilhører har små høyttalere "inne" i ørene. Dermed kan vi også betrakte lydfokusering med de hode-relaterte impulsresponsene.

Betrakt kildeplassering 2 i konteksten lydfokusering. Til tross for gode separasjonsverdier er det ikke synlig noe fokuspunkt i denne figuren. Det som imidlertid skjer, er at det *slippes gjennom mer energi i et annet punkt enn fokuspunktet*, etter endt inversfiltrering. Relativt til fokuspunktet slippes det igjennom 4.8dB mer i dette punktet. Dette vises bedre i figur 5.11, enn i figur 5.10. I figur 5.11 er L_2 -verdiene plottet mot den horisontale vinkelen ϕ , og elevasjonsvinkelen θ . Dette plottet gir et overblikk over kuleplottet. Punkter hvor det slippes igjennom mest energi er i området rundt punktet $(-60, 230)$.



Figur 5.11: Overblikk av separasjonen med hode-relaterte impulsresponser ved kildeplassering 2.

Betrakt igjen figur 5.8, som tilhører simuleringen med kildeoppsett 1. Vi legger merke til at området rundt fokuspunktet er "flekke", det er flere punkter som slipper igjennom like mye, eller mer energi enn fokuspunktet.

Dette kan sees tydeligere i figur 5.9, enn i figur 5.8. Det finnes et punkt hvor det slippes igjennom 1.6dB mer enn i fokuspunktet.

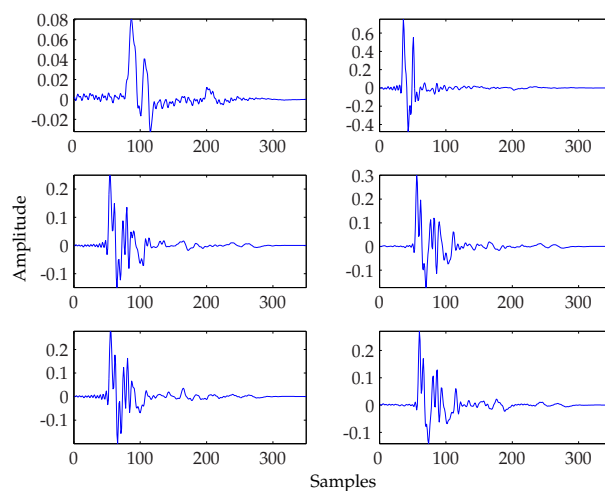
En slik ustabil fokuseringen så vi også i modellen i forrige kapittel. Dette var spesielt tilfellet for modeller med korte impulsrespons, se figur 4.10. Betyr dette at de hode-relaterte impulsresponsene ikke er tilstrekkelig innholdsrike? Vi undersøker derfor de HRIRene som inngår i matriseproblemet tilhørende systemoppsett 1, se figur 5.12. Impulsresponsenes plassering i plottet, tilsvarer plasseringen til de tilhørende konvolusjonsmatrisenes plassering i den sammensatte konvolusjonsmatrisen \mathbf{H} . Impulsresponsene har tilsammen 350 filtertapper, men vi ser at de har en "signifikant" del på cirka 40-60 tapper. Er det slik at det er den signifikante delen av impulsresponsene som her gjør seg gjeldende? På bakgrunn av resultatene i delkapittel 4.3 i forrige kapittel, sammenliknet med disse resultatene, kan det argumenteres at dette er tilfellet.

En obsevasjon

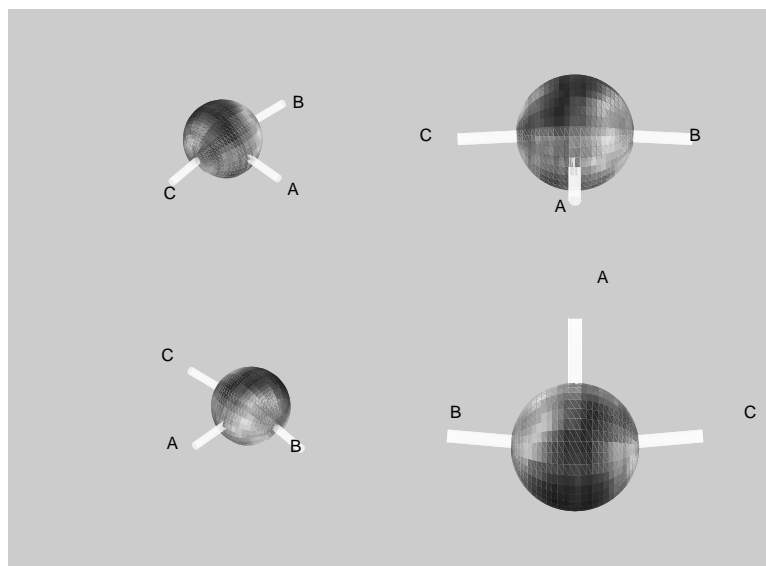
Når vi mennesker lytter til en ønsket kilde har vi en tendens til å snu hodet slik at denne kilden settes rett foran oss. Simuleringene utført i dette delkapitlet blir derfor kanskje litt kunstige.

Jeg gjennomfører en siste simulering med HRIRene, hvor jeg antar at det er en ønsket kilde rett foran testpersonen, i punktet (0,0), og to uønskede i punktene (-85,0) og (85,0). Resultatet av denne simuleringen kan sees i figur 5.13. Figuren viser *ingen synlig separasjon*, hvor $SIR_{ut} = 0.67\text{dB}$. Altså ingen separasjon oppnås, heller ingen fokusering, om vi ser på dette kildeoppsettet i den konteksten. På en annen side ville et slikt oppsett gi et vanskelig separasjonsproblem i og med at de uønskede kildene spilles av rett inn i ørene. Det er interessant å se at resultatene kan variere i så stor grad. Noen kildeplasseringer kan gi opptil 14.9dB forsterkning (jamfør kildeplassering 2), mens det ved andre kildeplasseringer ikke oppnås noen forsterkning.

Et naturlig skritt for videre forskning er så undersøke om hvorvidt det er mulig å kunne identifisere hvilke egenskaper ved impulsresponsene som gjør at det kan oppnås gode resultater.



Figur 5.12: Impulsresponser som inngår i den sammensatte konvolusjonsmatrisen \mathbf{H} ved kildeplassering 1.



Figur 5.13: “Hodeplott” med fokus inn på en kilde rett foran lytteren. Sylinderen merket A, skjærer kula der den ønskede kilden er. Sylindrene merket B og C skjærer kula der de uønskede kildene er.

Kapittel 6

Konklusjon og diskusjon

I denne masteroppgaven har jeg primært undersøkt det å fokusere lyd fra lydkilder til et punkt, og samtidig undertrykke andre punkter. Utgangspunktet for oppgaven var inspirert av menneskets evne til å separere ut en kilde blant en miks av flere. Dette ga opphav til noen spørsmål som ble stilt i problemstillingen. Med denne oppgaven har jeg forsøkt å besvare disse spørsmålene.

Simulasjonsmodellen

Ved simulering av lydfokusering så vi at ulikhet mellom impulsresponsene, mellom hver av lydkildene og punkter på kulen, var viktig for å få gode relative energinivåer og relative maksimale utslag mellom fokuspunkt og undertrykkelsespunkter. Denne typen rikhet i impulsresponsene bidro til god faktisk fokusering. Etterhvert som impulsresponsene ble likere hverandre, ble den faktiske fokuseringen forringet. Dersom impulsresponsene ble for like, kunne ingen fokusering forekomme. Vi så også at ulikhet mellom impulsresponsene gjorde at fokuspunktet ble mindre, og at dette ga en større frihet til å plassere undertrykkelsespunktene i forhold til fokuspunktet med tanke på de globale fokuseringsegenskapene til modellen.

Det ble også utført simuleringer for å undersøke hvorvidt lengre og mer innholdsrike impulsresponses ville forbedre fokuseringen. Disse simuleringene viste at innholdsrike impulsresponses ga bedre kontroll over lydfeltet som helhet. Med korte impulsresponses ble responsen på kulen veldig uforutsigbar, og vanskelig å kontrollere. Om man tar den faktiske fokuseringen med i betraktningen er det slik at mer innholdsrike impulsresponses ga relative energinivåer, relative maksimale utslag, mellom fokuspunkt og undertrykkelsespunkter, som lå nærmere gjennomsnittsverdiene til disse. Dette gjaldt også for tilpasningsfeilen knyttet til det inverse problemet.

Med kortere impulsresponser kunne det derimot forekomme relative energinivåer og tilpasningsfeil som var bedre enn ved lengre impulsresponser. Dette gjelder også i negativ retning, da verdiene også kan bli dårligere.

Ved fokusering med innholdsrike impulsresponser oppsto det også en jevn undertrykking av punkter som ikke inngikk i løsningen av det inverse problemet. Med en økning av antall lydkilder kunne den faktiske og den utilsiktede undertrykkingen gjøres bedre. Vi så indikasjoner på at innholdsrike impulsresponser også var viktige for modeller med flere lydkilder.

Den faktiske fokuseringen kunne bedres ved å gjøre om på forholdet mellom antall lydkilder og antall kontrollpunkter på kulen. Et oppsett med antall kontrollpunkter lik antall lydkilder ga de beste resultatene, med tanke på den faktiske fokuseringen. Ved å øke antall kontrollpunkter i forhold til antall lydkilder kunne den utilsiktede fokuseringen øke, samtidig som tilpasningsfeilen og den faktiske fokuseringen da ble dårligere. Man kunne foreta en avveining mellom faktisk fokusering og utilsiktet undertrykking. En avveining mellom faktisk fokusering og utilsiktet undertrykking kunne også gjøres ved velge lengden på de inverse filterne lenger enn impulsresponsene.

Et av spørsmålene jeg skulle forsøke å besvare med denne oppgaven var om lengre og mer innholdsrike impulsresponser ville gi bedre fokusering. På bakgrunn av resultatene fra kapittel 4 vil jeg argumentere for at dette er tilfellet.

Jeg skulle også forsøke å besvare spørsmålet om det var viktig med ulike tidsdomene-representasjon mellom impulsresponsene i rom. Dette har også vist seg å være viktig ved fokusering av lyd.

På spørsmålet om hvorvidt lydfokusering med innholdsrike impulsresponser kan utnyttes til å redusere antall lydkilder, til applikasjoner for hørbar lyd, må svaret bli nei. Det er de relative energinivåene mellom fokuspunkt og undertrykkelsespunktene som best forklarer hvor godt signalet høres i fokuspunktet relativt til undertrykkelsespunktet når musikk/tale skal fokuseres. Simuleringene viste at disse verdiene ikke ble høyere ved bruk av lengre impulsresponser.

De relative energinivåene, mellomfokuspunkt og undertrykkelsespunktene, er ved bruk av 2 lydkilder ikke gode nok til å gi tilstrekkelig hørbar forskjell når musikk/tale skal fokuseres. Ved bruk av 10 lydkilder begynner forskjellen å bli relativt god, for den faktiske fokuseringen. For å oppnå hørbar forskjell i de punktene som ikke inngår i den faktiske fokuseringen må antall lydkilder utvides ytterligere.

Eksperimenter på ekte data

Det ble også utført eksperimenter på ekte data. Impulsresponsen målt i et rom ble benyttet her. Jeg forsøkte å anskaffe innholdsrike impulsresponsen ved å måle disse gjennom et sett med reflektorer. Fokusering med disse impulsresponsene ble sammenliknet med fokusering med impulsresponsen målt uten reflektorene. Denne simuleringen, sammenliknet med resultater fra kapittel 4, ga indikasjoner på at reflektorene hadde ønsket effekt.

Ved bruk av impulsresponsene målt i et rom undersøkte jeg også hvor robust fokuseringen var mot støy. Disse testene viste at det var kun ved veldig dårlige signal-støy forhold at fokuseringen ble forringet. Ved bedre signal-støy forhold presterte løsningsfilterne mer tilfredstillende.

Til slutt utførte jeg noen forsøk der kildeseparasjon ble undersøkt. Kilder ble virtuelt plassert i posisjoner rundt hodet til en person. Ved noen kildeplasseringer kunne man oppnå gode separasjonsverdier. Ved andre kildeplasseringer kunne noe mindre gode resultater oppnås, og i noen tilfeller kunne ikke separasjon oppnås.

På samme måte som i kapittel 4 plottet vi L_2 kriteriet i alle punkter. Dette resulterte også her i et kule-plott. Vi betraktet også lydfokusering med hode-relaterte impulsresponsen. Dermed kunne resultatene fra dette delkapitlet lettere sammenliknes med resultatene fra kapittel 4. Det viste seg at lydfokusering med hode-relaterte impulsresponsen hadde likhetstrekk med fokusering med korte impulsresponsen i kapittel 4. Det så ut til at den "signifikante" delen av impulsresponsene var den som gjorde seg mest gjeldene.

Jeg forsøkte også å besvare spørsmålet om hvorvidt lengre og mer innholdsrike impulsresponsen ville bedre separasjon av kilder. På bakgrunn av resultatene i denne oppgaven er svaret på dette nei. SIR_{ut} kriteriet er omtrent det samme kriteriet som $f_2(.,.)$. $f_2(.,.)$ måler energi forholdet mellom to responsen, mens SIR_{ut} mellom energi forholdet mellom responsen som konvolveres med den ønskede kilden, og responsene som konvolveres med *alle* de uønskede kildene. Siden $f_2(.,.)$ ikke ga bedre verdier med lengre impulsresponsen vil heller ikke SIR_{ut} gi bedre verdier med lengre impulsresponsen.

Hvorvidt ørets tilsynelatende kaotiske form hjelper menneskene til å separere kilder fra hverandre er fremdeles ett åpent spørsmål. Dette spørsmålet har jeg ikke klart å besvare med den matematiske modellen som jeg har benyttet. Riktig nok har forsøkene gitt spennende resultater som det er verdt å undersøke ved videre forskning.

6.1 Videre forskning

Denne oppgaven har gitt mange interessante resultater. Det er fremdeles ting som bør forskes videre på. Nedenfor har jeg listet opp ting jeg mener er viktige for videre arbeid.

- Implementasjon av system med flere høyttalere i et rom. Man bør bruke flere og bedre reflektorer. Impulsresponsene bør måles i et ekkofritt rom.
- Det bør undersøkes hvorvidt det er mulig å fokusere lyden med impulsresponsene tilhørende det “lokale området”, altså impulsresponsene målt gjennom et sett med reflektorer, i et rom som ikke er lyddempet. Dersom dette er tilfellet vil man kunne minke kompleksiteten på beregningene av de inverse filterne betraktelig.
- Gå mer inn i matrisens struktur for å kunne bestemme optimal avveining mellom faktisk fokusering og utilsiktet undertrykking.
- Undersøke kildeparasjon med HRIR som er tatt opp i rom med mer ekko enn et lyddempet rom. Det er interessant å undersøke om hvorvidt dette vil stabilisere separasjonsverdiene, altså at det gis jevnere separasjon ved forskjellige kildeplasseringer.
- Gå ennå mer i dybden ved kildeparasjon med hode-relaterte impulsresponses. Undersøke hvorvidt det er mulig å indentifisere hva som er årsaken til det ved noen kildeplasseringer gis god separasjon.

6.2 Applikasjoner

Det å kunne fokusere lyd med høyttalere kan resultere i mange interessante applikasjoner.

- “Annoying Speaker Removal”. Et fjernstyrt høreapparat, der man ved hjelp av en fjernkontroll kan bestemme posisjonene til kilden man vil høre på, og kildene man ikke vil høre på. Her kan man også få muligheten til å øke antall sensorer. Ulempen er da at utstrekningen til høre-apparatet også vil øke.
- Lydfokusering på f.eks flyplasser hvor man kan fokusere lyd inn i soner hvor viktig informasjon gis. Ved bruk av skype-telefoni kunne man gis muligheten til å snakke i telefon på café, uten hode-telefoner, uten å forstyrre personer i nærheten, ved å fokusere lyden høyttaleren mot personen foran datamaskinen.

- Overvåkning, hvor man i tillegg til kamera kan benytte seg av en antenne med mikrofoner for å fokusere lytteområdet til det samme området som kameraet viser.
- Lydfokusering i lyddempede rom.
- Kunst-installasjoner.

Tillegg A

Impulsresponsmålinger

Impulsrespons er en god måte å representere lineære systemer på. Den gir rom for flere måter å analysere systemegenskapene: Dersom man transformerer impulsresponsen til z-rommet, vil man kunne bestemme egenskaper som stabilitet, kausalitet, om systemet er minimum/maksimum eller blandet fase[32]. Med en transformasjon til frekvensdomenet vil man kunne analysere hva slags frekvensinnhold systemet har. Hvordan skal man anskaffe disse impulsresponsene, og hva slags signaler bør brukes til dette formålet? Impulsresponsen kan bestemmes av et inngangs-utgangs forhold[32]. Dette forklares nærmere her.

Hvis man antar et rommets propageringsegenskaper, gitt av ekkoer og gjenklang, som et lineært system, kan man modellere transformasjonen av et signal som propagerer fra et punkt i rommet til et annet, med konvolusjonsoperatoren.[9]. For å estimere impulsresponsen kan man gå fram slik som ble indikert i seksjon 2.5. Man må anskaffe et passende målesignal. Ta målinger av systemet, og sammenlikne målesignalet og målingene av systemet. Det mest naturlige valget av målesignalt vil kanskje være å rett og slett bruke et impuls-signal. Den målte responsen vil da være lik impulsresponsen[23]. Teoretisk sett inneholder en impuls alle mulige frekvenser, dette kan sees ved å ta den diskrete Fourier-transformen til et slikt signal: Anta $x(n) = \delta(n)$, den diskrete Fourier transformen til $x(n)$ er

$$X(\omega) = \mathcal{DFT}(x(n)) = \sum_{-\infty}^{\infty} x(n)e^{-j\omega n} = \sum_{-\infty}^{\infty} \delta(n)e^{-j\omega n} = 1,$$

siden

$$\delta(n) = \begin{cases} 1 & \text{hvis } n = 0 \\ 0 & \text{ellers} \end{cases}$$

Frekvensresponsen til et slikt signal er flat. Om man anvender dette

signalet vil man teoretisk sett kunne få kjennskap til karakteristikene til systemet over alle frekvenser. Siden konvolusjon i tidsdomenet er det samme som multiplikasjon i frekvensdomenet[32]

$$Y(\omega) = \mathcal{FFT}\{x(n) * h(n)\} = H(\omega)X(\omega) = H(\omega).$$

Riktignok innehar dette signalet teoretisk sett de egenskapene som gjør dette til det ideelle målesignalet. I praksis vil et impuls-signal være båndbegrenset. Slike signaler er et godt valg om signalbanen er rent elektrisk, og det ikke er en akustisk signalbane[27]. I følge ISO-3382 standarden skal målesignaler oppfylle følgende kriterier[23]

- Det bør være mulig å kunne sende signalet i alle retninger.
- Det bør ha en tilstrekkelig dynamisk bredde for å unngå at signalet ødelegges av bakgrunnsstøy.
- Det bør være repeterbart.

Historisk sett finnes det mange metoder for å måle impulsresponsen til et rom på[27]. En metode som er veldig mye brukt, spesielt innen akustikk, er målinger med såkalt maksimum lengde sekvenser (MLS), også kalt pseudo-tilfeldige sekvenser. Denne metoden for å måle akustiske impulseresponser ble først foreslått av Schröder i 1979[23]. Maksimum-lengde-sekvenser er binære sekvenser med lengde $K = 2^m - 1$. Disse sekvensene har noen egenskaper som lenge gjorde at de var det foretrukne til dette formålet. Den viktigste av disse egenskapene er at den periodiske autokorrelasjonsfunksjonen har en spesiell form. Dette gjorde at ekstraheringen av den akustiske impulsresponsen lot seg gjennomføre raskt.

A.1 Impulsresponsmålinger med sinus-sveip

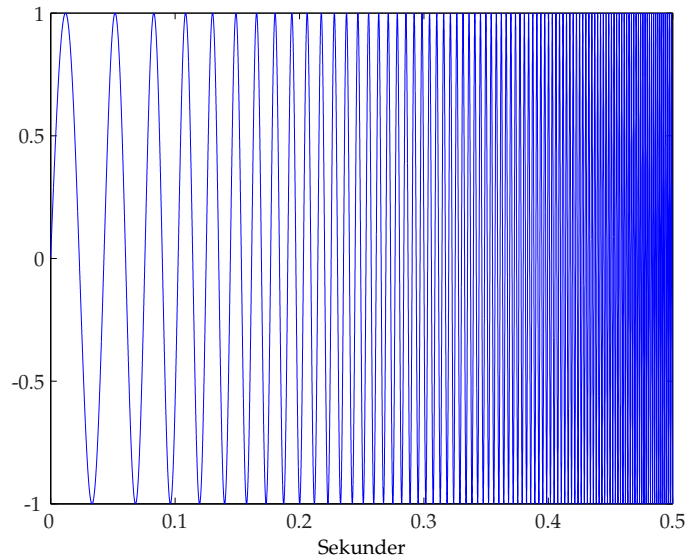
MLS signaler gir matematisk sett en elegant metode å måle impulsresponser på, med de byr imidlertid på noen praktiske problemer[27]. For å måle impulsresponser med MLS metoden antar man at systemet er perfekt lineært. Når denne antagelsen ikke er helt sann kommer det til syne tidsaliasing-effekter i den målte impulsresponsen[9]. Disse effektene sees som forvrengningstopper i impulsresponsen og kan minne om alisasings topper der den første delen av impulsresponsen er foldet inn i den egentlige impulsresponsen[9]. I tillegg til dette krever impulsresponsmålinger med MLS metoden at inngang og utgangssignalet er tett synkronisert for å få fornuftige resultater.

Det finnes en metode som ikke har disse begrensningene. Metoden ble først foreslått av Angelo Farina i 2000. Det viser seg at denne måten å måle impulsresponser på har noen fordeler i forhold til MLS metoden.

Det som kanskje er den fremste fordel er at metoden klarer å skille ut de harmoniske forvrengningene i dekonvolusjonen. [17] sier: *...Harmonisk forvrengning er en type forvrengning av signalet, hvor det opprinnelige signalet tilføres harmoniske overtoner for hver fordobling av inngangssignalets frekvens. Harmonisk forvrengning betegnes også som amplitudeforvrengning eller ulineær forvrengning.* Med en denne metoden kan man ekstrahere den lineære impulsresponsen[9]. Forskning har vist at signal-støy forholdet er bedre[27] enn med MLS-metoden. Impulsresponsen kan i tillegg beregnes fra de råe målingene uten noen form for synkronisering av dataene[23]. Metoden baserer seg på et sinus-signal med eksponensielt økende frekvens som målesignal[9]

$$x(t) = \sin\left[\frac{T\omega_1}{\ln(\frac{\omega_2}{\omega_1})} e^{\frac{t}{T}\ln(\frac{\omega_2}{\omega_1})-1}\right]. \quad (\text{A.1})$$

ω_1 er startfrekvensen, ω_2 er slutfrekvensen, intervallet mellom ω_1 og ω_2 er frekvensintervallet vi får kjennskap til systemet over. T er varigheten til signalet i sekunder. Man kan også benytte seg av et sinus-signal der frekvensen øker lineært. Man vil da ikke skille ut de harmoniske forvrengningene, men metoden vil fortsatt være mer robust mot aliasing[9]. Ekstraheringen av impulsresponsen kan utføres ved at man lager et inversfilter i henhold til Farina, detaljer om hvordan man konstruerer et inverst filter til dette målesignalet finnes i [35]. Målingen av $x(t)$ konvolveres så med dette inverse filteret.



Figur A.1: Sinussignal med eksponensielt økende frekvens, 0.5 sekunder av et målesignal på 1 sekund

Tillegg B

NORSIG 2006 artikkelbidrag

Underveis i arbeidet med denne masteroppgaven, ble det skrevet en artikkel som tok for seg noen aspekter av denne masteroppgaven. Resultatet ble publisert på "NORSIG 2006 7th NORDIC SIGNAL PROCESSING SYMPOSIUM" Vedlagt er artikkelbidraget til denne konferansen

Source Enhancement and Suppression with Binaural Head Related Impulse Response Functions

Tobias Dahl, Bjørn Cato Syversrud, Sverre Holm

Department of Informatics, University of Oslo
P. O. Box 1080, Gaustadallen 2, Blindern
Oslo, NORWAY

ABSTRACT

Within the field of wireless communications, it has been known for a while that "echoes" in the channels can be used constructively to improve communication success factors, and not only as a destructive feature as in the case of inter-symbol interference. This is particularly true for so called MIMO (Multiple Input Multiple Output) systems. In this paper, we propose to use echoes constructively for the purpose of improving audio source separation and suppression of signals from unwanted directions. Whereas this problem has long been addressed within acoustic (spatial) array technology, this paper shows that the temporal dimension plays an important role. The different signal echoes arising from different direction-dependent paths that the signals may travel, can be used to enhance the spatial array effects.

1. INTRODUCTION

The so-called "cocktail party effect" is the basis of hundreds of scientific papers. A large number of algorithms have been proposed to deal with source separation in a large number of settings, many of them aiming at reproducing the human ability to separate out single speakers in mixes of numerous voices. For a recent overview, see [1].

At the same time, the Head Related Transfer Function (HRTF) is being examined by others. Some suggest using them for improving audio experiences by "coloring" the sound according to its direction of arrival [2] while others relate the HRTFs to the cocktail party effect [3]. Although it is generally understood that the HRTFs play an important role in speech separation, it is still an open question just how much, and just how the direction-specific coloring of the speech sources help the listener in the separation process.

In this paper, we suggest a conceptually simple method for employing HRTF measurements, via the equivalent *Head Related Impulse Responses* (HRIR), for the purpose of source

separation. Successful source separation relies on two different still interrelated tasks to be solved: The first is to enhance the signals of the desired source, the second to simultaneously suppress the signals of the other sources.

The proposed method is based on a simple criterion, and shows promising separation qualities. Bringing the HRIRs themselves into the equation, one might also ask if not many of today's attempts at solving the cocktail party problem are indeed ill-posed. The cocktail party effect was first described in psychology [4], where the existence of two human ears was probably taken for granted. It could then well be that attempts at separating sources merely by using a setup of omni-directional microphones rules out a crucial element in the recipe for successful separation - the shape of the outer ear, the pinna.

Our results are related to those of [5] and [6] where the reverberations characteristics of the room acoustics is said to add super-resolution characteristics to the temporal and spatial focusing, although their work deals with the global characteristics of a room whereas our work deals with the local environment of the head and the outer ear.

From our results, it can be argued that the longer, more complex and more direction-dependent the echoes are, the better the separation. This leads us to the hypothesis that the "chaotic" shape of the human ear is an important factor in enabling humans to separate directional signals so well.

2. METHOD

Assume that a signal $x(t)$, transmitted from a specific point in space, is received into the human ear. Let $h_1(\cdot)$ denote the K -tap FIR filter corresponding to the HRIR,

$$y(t) = h_1(t) * x(t) = \sum_{i=0}^{K-1} x(t-i)h_1(i) \quad (1)$$

Assume $x(t)$ to be a signal of length N . Then, the above equation can be written as

$$\mathbf{y}(t) = \mathbf{H}_1 \mathbf{x}(t) \quad (2)$$

where $\mathbf{H}_1 \in R^{N+K,N}$ is a convolution matrix containing sub diagonal filter taps which are the elements in the filter $h_1(\cdot)$. In order to recover the signal $x(t)$ successfully, it is of interest to find an inverse filter. There are many ways of doing this, but one is to solve the problem

$$\min \arg_{\mathbf{g}_1} \|\mathbf{H}_1 \mathbf{g}_1 - \mathbf{e}\|_2^2 \quad (3)$$

where \mathbf{g}_1 takes the place of the inverse filter, and \mathbf{e} is vector with zeros everywhere except a one in the middle coefficient. Solving this problem produces the ideal inversion filter $g_1(\cdot)$ in the least squares sense. A critical factor in producing the inverse filter is the inverse filter length, which is decided by the dimensions of the matrix \mathbf{H}_1 . Intuitively, this dictates “how long one should listen” before trying to invert the process. Assuming that the filter taps in $g_1(\cdot)$, taken as the elements of the vector \mathbf{g}_1 have been computed, a future signal source $x(t)$, received as $y(t)$ after having passed through the filter $h_1(\cdot)$ as $y(t)$, can be successfully reconstructed as

$$\hat{x}(t) = g(t) * y(t) \quad (4)$$

2.1. Two sensors are better than one

Returning to Eq. 3, let us consider the equation from a linear model point of view. The minimization criterion could be associated with a linear model,

$$\mathbf{H}_1 \mathbf{g}_1 = \mathbf{e} + \epsilon \quad (5)$$

Finding the inverse filter is then just a matter of minimizing the L_2 norm of the residual $\epsilon = \mathbf{H}_1 \mathbf{g}_1 - \mathbf{e}$. A natural question is what is needed for the error level to become zero or close to zero. To obtain zero error, one must linearly combine the columns of the matrix \mathbf{H}_1 with the weights held in \mathbf{g}_1 such that the combination becomes identical to \mathbf{e} . This is the same as to say that \mathbf{e} is in the column space of \mathbf{H}_1 . This property will not hold in a number of different cases, such as e.g. the filter $h_1(\cdot)$ being bandpass. The best possible inversion then corresponds to combining the columns of \mathbf{H}_1 to become the band limited version of the Dirac impulse response, the perfect one here represented by the unit element in the middle of \mathbf{e} . However, as shown in [7], it is possible to do better if *two independent recordings* of the signal $x(t)$, observed through different filters $h_1(\cdot)$ and $h_2(\cdot)$ are available,

$$y_1(t) = h_1(t) * x(t) \quad (6)$$

$$y_2(t) = h_2(t) * x(t) \quad (7)$$

Seeking to recover a future $x(t)$ from future observed signals $y_1(t)$ and $y_2(t)$ can be accomplished by seeking a *combined* inverse filter, e.g filters $g_1(t)$ and $g_2(t)$ such that

$$h_1(t) * g_1(t) + h_2(t) * g_2(t) = \delta(t) \quad (8)$$

where $\delta(t)$ is the Dirac delta function. If this property holds, then using the filters g_1 and g_2 on $y_1(t)$ and $y_2(t)$ and summing the contributions, one gets

$$\begin{aligned} & y_1(t) * g_1(t) + y_2(t) * g_2(t) \\ &= h_1(t) * x(t) * g_1(t) + h_2(t) * x(t) * g_2(t) \\ &= x(t) * [h_1(t)g_1(t) + h_2(t)g_2(t)] \\ &= x(t) * \delta(t) = x(t) \end{aligned} \quad (9)$$

In matrix/vector notation, (8) is:

$$\mathbf{H}_1 \mathbf{g}_1 + \mathbf{H}_2 \mathbf{g}_2 = \mathbf{e} + \epsilon \quad (10)$$

or, in block notation

$$(\mathbf{H}_1 \ \mathbf{H}_2) \begin{pmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{pmatrix} = (\mathbf{H}_1 \ \mathbf{H}_2) \mathbf{g} = \mathbf{e} + \epsilon \quad (11)$$

where $\mathbf{g} = [\mathbf{g}_1^T \mathbf{g}_2^T]^T$. Minimizing the norm of the residual ϵ in the L_2 sense is then equivalent to solving

$$\min \arg_{\mathbf{g}} \|(\mathbf{H}_1 \ \mathbf{H}_2) \mathbf{g} - \mathbf{e}\|_2^2 \quad (12)$$

The filter taps of $g_1(\cdot)$ and $g_2(\cdot)$ held in the first and second half of the vector \mathbf{g} respectively, can then be used for the purpose of combined/additive inversion as sketched in (8).

2.2. Suppression of unwanted sources/directions

Assume now a slightly different problem where a signal is transmitted from a point in space, arriving at the two sensors through the filters $h_3(\cdot)$ and $h_4(\cdot)$ respectively. This time, we don't want to find an inverse filter, but rather a suppression filter, so that, when the two filtered versions of the signals are combined after reception, they amount to a zero signal. Moving directly to the matrix form described above, this could be accomplished by seeking a vector $\mathbf{g} = [\mathbf{g}_1^T \mathbf{g}_2^T]^T$ such that

$$(\mathbf{H}_3 \ \mathbf{H}_4) \begin{pmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{pmatrix} = (\mathbf{H}_3 \ \mathbf{H}_4) \mathbf{g} = \mathbf{0} \quad (13)$$

The corresponding least squares problem is

$$\min \arg_{\mathbf{g} \neq \mathbf{0}} \|(\mathbf{H}_3 \ \mathbf{H}_4) \mathbf{g}\|_2^2 = \min \arg_{\mathbf{g} \neq \mathbf{0}} \|(\mathbf{H}_3 \ \mathbf{H}_4) \mathbf{g} - \mathbf{0}\|_2^2 \quad (14)$$

where $\mathbf{0}$ is the vector of zero elements matching the number of rows in \mathbf{H}_3 and \mathbf{H}_4 . The constraint $\mathbf{g} \neq \mathbf{0}$ is needed to avoid the trivial all-suppressing solution.

2.3. Combined detection/inversion and suppression

The efforts made in the two proceeding sessions will now be combined. We want to detect the signal transmitted and received through the filters $h_1(\cdot)$ and $h_2(\cdot)$, and simultaneously, to suppress the signal transmitted and received through the filters $h_3(\cdot)$ and $h_4(\cdot)$. We are assuming that the sensors are not simply equidirectional/point sensors, but that they have directive sensitivity and diversity: Depending on the direction of arrival into the two sensors, the signal is filtered in a different manner. This could be due to the shape of the outer ear (the pinna). We then want to solve two problems simultaneously, (14) and (12), for the same choice of \mathbf{g} : The filter taps of $g_1(\cdot), g_2(\cdot)$ held in \mathbf{g} , should, ideally, simultaneously work to enhance the signal passing through h_1, h_2 while suppressing the signal passing through h_3, h_4 . Assuming all filters to be contained in sub diagonal filter matrices $\mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3, \mathbf{H}_4$ with the same number of columns, this can be expressed in block form as

$$\min \arg \mathbf{g} \left\| \begin{pmatrix} \mathbf{H}_1 & \mathbf{H}_2 \\ \mathbf{H}_3 & \mathbf{H}_4 \end{pmatrix} \mathbf{g} - \begin{pmatrix} \mathbf{e} \\ \mathbf{0} \end{pmatrix} \right\|_2^2 \quad (15)$$

Assume further that signals from another direction, that would pass through the filters h_5 and h_6 are to be suppressed as well. The above optimization problem could then be expanded to

$$\begin{pmatrix} \mathbf{H}_1 & \mathbf{H}_2 \\ \mathbf{H}_3 & \mathbf{H}_4 \\ \mathbf{H}_5 & \mathbf{H}_6 \end{pmatrix} \mathbf{g} = \begin{pmatrix} \mathbf{e} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} + \epsilon \quad (16)$$

Letting \mathbf{H} denote the matrix of filter sub blocks on the left hand side, the performance of the combined detection and suppression could be assessed by considering the prediction $\mathbf{y} = \mathbf{H}\mathbf{g}$. The vector $\mathbf{y} \in R^Q$ should, ideally, have a unit element in the middle of the top sub vector, here denoted by the vector element index P , and zero everywhere else.

2.4. What makes separation successful?

Consider now the situation where the filters h_1, \dots, h_6 only has one tap each. It can then be shown that the corresponding filter matrices $\mathbf{H}_1, \dots, \mathbf{H}_6$ are just multiples of the identity matrix. But in this case, no separation can be obtained. If the model is

$$\begin{pmatrix} h_1 \cdot \mathbf{I} & h_2 \cdot \mathbf{I} \\ h_3 \cdot \mathbf{I} & h_4 \cdot \mathbf{I} \\ h_5 \cdot \mathbf{I} & h_6 \cdot \mathbf{I} \end{pmatrix} \mathbf{g} = \begin{pmatrix} \mathbf{e} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} + \epsilon \quad (17)$$

then there is no way a single vector \mathbf{g} could obtain the desired result. The sub vectors in \mathbf{y} would all be very similar, no distinction could be made that would make the first sub vector contain a unit element in the middle, and the successive sub vectors being zero at the same time. The likely outcome of such a procedure would be a poor compromise between separating out the first source and cancelling

out the others, none of the tasks done in a very satisfactory manner. The same would be the case if all the filters h_1, \dots, h_6 were similar or the same. This leads us to the insight that the crux for successful separation is *diversity* in the filters. Furthermore, the matrix dimensions are of importance, and the filter lengths relative to those dimensions. A general remark would be that the longer (and more different) the filters h_1, \dots, h_6 are, the better the separation/suppression process is.

3. RESULTS

3.1. Data set

To illustrate separation and enhancement performance using real Head Related Impulse Responses, several simulations were performed. The measurements used are the Head Related Impulse Responses (HRIRs), available from [8]. Their data was measured at the entrance of the left and right ear canal. Several loudspeakers were mounted on a semi-circular, horizontal rotating loop to cover all azimuths and elevations and the HRTF was measured at 1093 points. Data collection was performed in a sound attenuating room with miniature microphones inserted into the subject's ears. The measurements have then been filtered and sampled at 83.3k samples/sec.

3.2. Simulation

In the example that follows, three source directions were present, one desired, and two interferers. In each instance, the three pairs of HRIR filters were extracted from the database, and used as the filter pairs defining the sub matrices $\mathbf{H}_1, \dots, \mathbf{H}_6$. Once the vector \mathbf{g} has been calculated, we are equipped with two filters $g_1(\cdot), g_2(\cdot)$ which, under very ideal circumstances, should remove the signals coming from the interference directions. If we were even more lucky, it would suppress sound from *all other directions* except the desired direction also. To get an idea of the degree of success in the latter assignment, we can visualize the angular extraction/suppression performance on a sphere.

We computed the maximum value of the response in each direction:

$$L(\phi_i, \theta_i) = \|\mathbf{H}_1(\phi_i, \theta_i) \mathbf{H}_2(\phi_i, \theta_i) \mathbf{g}\|_\infty \quad (18)$$

based on the HRIRs, $\mathbf{H}_1(\phi_i, \theta_i)$ and $\mathbf{H}_2(\phi_i, \theta_i)$, available in the dataset. We then plotted this value on the (fixed-radius) sphere as an intensity value.

In Fig. 1, we show a desired source A which was about right outside the right-hand ear at azimuth, elevation $(85^\circ, 0^\circ)$, and at the same time interfering sources at azimuth, elevation $(0^\circ, 0^\circ)$ B, and $(-85^\circ, 0^\circ)$ (C, outside the left-hand ear) were present. With one $L(\phi_i, \theta_i)$ intensity value per spatial direction, we then made a spherical plot where the

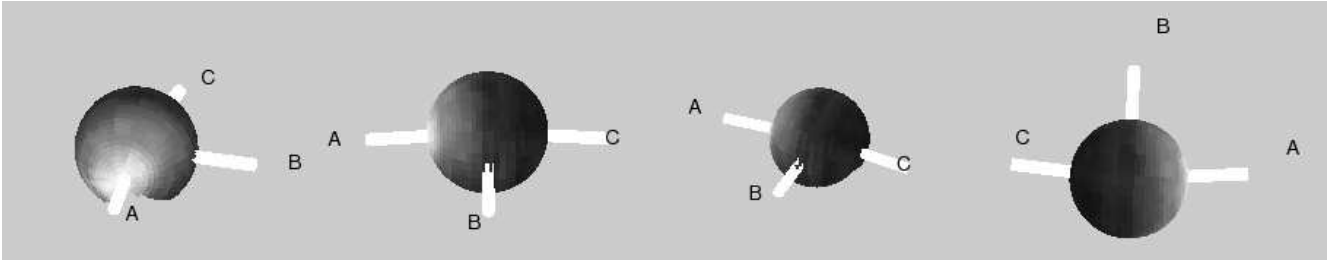


Fig. 1. Source separation of a desired source, A, in the presence of interfering sources, B and C. Bright patches on the sphere corresponds with high signal throughput, dark patches to low throughputs. The same plot is shown from four different angles, and the gap on the sphere corresponds area for which no HRIRs are available.

intensity of the patches correspond to the maximum signal strength in that direction. A brighter patch means a stronger signal. Since the HRIR library is unevenly sampled in space, bilinear interpolation of the $L(\phi_i, \theta_i)$ over the sphere was used to increase the number of colored patches and improve visibility. The direction of arrival for the three sources are marked by cylinders penetrating the sphere, the main axis of the cylinder being parallel with the direction of the sources.

3.3. Results

For the source of interest ($85^\circ, 0^\circ$), the relative maximum signal strength L was 15.5 dB relative to the source at ($0^\circ, 0^\circ$), and 19.9 dB relative to the source at ($-85^\circ, 0^\circ$). This is encouraging, especially if one takes into account that the source which played more or less straight into one ear is suppressed almost completely in order to focus in on the sound coming into the other ear.

The results were more or less unchanged if only the 36 most central samples of the impulse responses, corresponding to 0.43 ms or about 14.9 cm travel time in air were used rather than the original 350 samples of the HRIRs of the database.

4. CONCLUSION

These initial tests indicate that a time domain interpretation of sound traveling around the head and through the outer ear is important for understanding how our ability to enhance and suppress sound sources works.

In the present paper, a very simple linear model was used for separation, and more advanced model are readily available. In particular, when dealing with noise in the impulse response estimates, better inverse filters may be obtained by minimizing a different criterion than the least squares, e.g. by introducing noise-robust regularization for the purpose of minimizing the expected mean square error in the presence of noise. This will also serve to robustify the

matrix inversion with respect to small eigenvalues in the \mathbf{H} matrix. Another aspect that needs to be studied is how well the norm in Eq. (18) represents the actual performance of the auditory system, and if other norms are more relevant.

A natural next step would also be to examine separation performance for bandlimited signals, and to match these findings with empirical knowledge of the human perception systems ability to perform separation of high and low tones. Studies of separation in various planes related to the head and body is yet another natural step.

5. REFERENCES

- [1] S. Rickard, P. O'Grady, and B. Pearlmutter, "Survey of sparse and non-sparse methods in source separation," *Int. J. Imag. Syst. Techn.*, 2005.
- [2] M. Dempsey, "Method of customizing HRTF to improve the audio experience through a series of test sounds," *US Patent 6768798*, 1997.
- [3] D. Brungart and B. Simpson, "Distance-based speech segregation in near-field virtual audio displays," in *Proc. Int. Conf. on Auditory Display*, pp. 169–174, 2001.
- [4] E. C. Cherry, "Some experiments on the recognition of speech, with one and two ears," *J. Acoust. Soc. Am.*, vol. 25, pp. 975–979, 1953.
- [5] S. Yon, M. Tanter, and M. Fink, "Sound focusing in rooms. The time-reversal approach," *J. Acoust. Soc. Am.*, vol. 113, no. 3, pp. 1533–1543, 2003.
- [6] S. Yon, M. Tanter, and M. Fink, "Sound focusing in rooms. II. The spatio-temporal inverse filter," *J. Acoust. Soc. Am.*, vol. 114, no. 6, pp. 3044–3052, 2003.
- [7] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Acoust., Speech, Sign. Proc.*, vol. 36, no. 2, 1988.
- [8] E. Grassi, J. Tulsi, and S. Shamma, "Measurement of head related transfer functions based on the empirical transfer function estimate," in *Proc. Int. Conf. on Auditory Displays*, (Boston, MA, USA), pp. 119–122, 2003.

Tillegg C

Program-beskrivelser

Her er vedlagt utvalgt programbeskrivelser for de viktigste programmene benyttet ved simuleringene, simulasjonene på HRIR-databasen, og for de virkelige opptakene.

Disse funksjonen og skriptene brukes til simuleringskapitlet.

- **lagDistribusjonerAvSepVerdierOgFeil.m:**

Skript som trekker tilfeldige filtere og gjør 4 simuleringer ved å bruke 10, 100, 200 og 500 samples av de tilfeldige filterne. Det lages 4 sammensatte konvolusjons matriser for simuleringer av 2 lydkilder, ett fokuspunkt og 2 undertrykkelsespunkter. Det beregnes relative energiverdier $f_2(.,.)$ og $f_\infty(.,.)$ mellom fokuspunktet og begge undertrykkelsespunktene, samt tilpassningsfeilen ϵ . Dette gjøres 1000 ganger. Distribusjonene av separasjonsverdier målt med 2-norm, inf-norm og tilpassningsfeilen lagres, og gjennomsnittsverdier beregnes..

- **dynamiskFokuseringsModell.m:**

Funksjon som tar imot en celle-matrise S hvor hver forekomst i S er en matrise med modellerte impulsresponser, antall impulsresponser pr. forekomst i S er avhengig av hvor mange lydkilder som modelleres. Impulsresponsene i S kan være glattet i henhold til glattingsalgoritmen i 3.3. Den tar også imot hvilke punkter som er fokuspunkt og undertrykkelsespunkter, hvor lange impulsresponser som skal brukes, hvor lange de inverse filterne skal være og antall lydkilder som modelleres. Funksjonen plukker ut impulsresponsene tilhørende fokuspunktet og undertrykkelsespunktet og danner den sammensatte konvolusjonsmatrisen H . Deretter lages responsvektoren d , og likningssystemet løses. Løsningsvektoren g deles deretter opp i M de filterne. Separasjonsverdiene for den faktiske fokuseringen beregnes, samt tilpassningsfeilen og disse skrives til skjerm. Til slutt løpes det

igjennom alle forekomster i S hvor det beregnes L_2 og L_∞ i alle tilgjengelige punkter, verdiene L_2 og L_∞ lagres så i hver sin matrise for deretter å bli plottet på en kule.

- **randomImpulsresponseSphere.m:**

Funksjon tar imot en impulsresponslengde og hvor mange punkter på kulen som skal modelleres. Den konstruerer deretter en cellematrise S med M impulsreponser av lengde N til hvert punkt. Impulsresponsene modelleres som uniformt distribuerte tall i intervallet $[-1,1]$. Funksjonen returnerer matrisen S og dimensjonene til S .

- **spaceAveragedSphere.m:**

Funksjon glatter matrisen S i henhold til glattingsalgoritmen i 3.3. Og returnerer den glattede matrisen S_{avg} .

Disse funksjonene brukes til simuleringene av de hode-relaterte impulsresponsene.

- **HRTF_sphere.m:**

Funksjon som tar imot koordinatene til fokuspunktet og de 2 undertrykkelsespunktene. Impulsresponsene tilhørende disse punktene ekstraheres fra HRIR-databasen. Den sammensatte konvolusjonsmatrisen H , og responsvektoren \underline{d} konstrueres, og likningsystemet 2.21 løses. Løsningen \underline{g} inneholder de to optimale filterne $g_1(n)$ og $g_2(n)$. Deretter løpes det igjennom alle tilgjengelige punkter hvor de inverse filterne benyttes på impulsreponsene tilhørende disse punktene for å beregne L_2 for hvert punkt. L_2 til hvert tilhørende vinkelpar lagres i en matrise I sammen med vinkelparet. Til slutt beregnes SIR_{ut} som skrives til skjerm.

- **matlab_kule.m:**

Denne funksjonen kalles etter HRTF_sphere.m og lager kule-plottene tilhørende simuleringene med HRIR. Funksjonen går først igjennom alle vinklene og plotter sylindrene som skjærer kule-plottet. Deretter går den på nytt igjennom alle vinklene (ϕ_i, θ_i) og setter L_2 verdiene inn på posisjoner i en matrise C som passer overens med hvordan matlab plotter en kule og å utfører en bi-lineær interpolasjon i de punktene hvor det ikke finnes noen verdier. Dette programmet er skrevet av Tobias Dahl.

De siste funksjonene er brukt til impulsresponsmålinger og for simuleringer med impulsresponsene målt i rom.

- **impulseResponseGrid:**

Dette skriptet genererer et sinussignal med eksponentielt økende frekvens (målesignalet), samt et inverst filter til dette signalet. Hvor

mange punkter som som det skal måles impulsresponser i gis lokalt i programmet. Ved hjelp av programpakken SND spilles målesignalet av og målinger kan gjøres med en mikrofon i rommet. Deretter dekonvolveres impulsresponsene ut av opptakssignalet ved at opptakssignalet konvolveres med det inverse filteret til målesignalet. Impulsresponsene lagres så i en matrise.

- **fokuseringMedStoy:**

Et skript som først trunkerer de målte impulsresponsene, deretter settes konvolusjonsmatrisene tilhørende fokuspunkt og undertrykkel-sespunktene settes sammen til matrisen \mathbf{H} , vektoren $\underline{\mathbf{d}}$ konstrueres og likningssystemet 2.21 løses. Løsningsvektoren \mathbf{g} deles opp i filterne $g_1(n)$ og $g_2(n)$. Deretter genereres det støy sekvenser i henhold til de valgte signal-støy forholdene betraktet i denne oppgaven. Disse støy sekvensene legges så til impulsresponsene tilhørende de 15 punktene målt i rommet. Deretter beregnes L_2 verdiene i alle punktene ved de forskjellige signal-støy forholdene. Resultatet av dette lagres i en matrise som deretter plottes.

Bibliografi

- [1] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano. The cipic hrtf database. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001.
- [2] V. Ralph Algazi, Richard O. Duda, Ramani Duraiswami, Nail A. Gumerov, and Zhihui Tang. Approximating the head-related transfer function using simple geometric models of the head and torso. *The Journal Of The Acoustical Society of America*, 112 (5), Pt. 1, Nov., 2002.
- [3] Shoko Araki, Shoji Makino, Yoichi Hinamoto, Ryo Mukai, Tsuyoki Mishikawa, and Hiroshi Saruwatari. Equivalence between frequency-domain blind source separation and frequency domain adaptive beamforming for convolutive mixtures. *EURASIP Journal on Applied Signal Processing* 2003:11,1157-1166, 2003.
- [4] Peter Blomgren, George Papanicolaou, and Hongkai Zhao. Super-resolution in time-reversal acoustics. *The Journal of the acoustical society of america*, January, Volum 111, pp. 230-248, 2002.
- [5] Laurent Couvreur. Blind deconvolution for multi-microphone speech dereverberation: Application to asr in reverberant environments. *Proc. 3rd IEEE Benelux Signal Processing Symposium (SPS-2002), Leuven, Belgium, March 21-22, 2001*.
- [6] Tobias Dahl, Bjørn Cato Syversrud, and Sverre Holm. Source enhancement and suppression with binaural head related impulse response functions. *NORSIG*, 2006.
- [7] Ronald D. Degroat and Eroc M. Dowling. The data least squares problem and channel equalization. *IEEE Transactions on Signal Processing*, VOL41. NO. 1, January, 1993.
- [8] Carsten Draeger, Jean Christian Aime, and Mathias Fink. One-channel time-reversal in chaotic cavities: Experimental results. *The Journal of the Acoustical Society of America*, 105(2), Pt. 1, February, 1999.

- [9] Angelo Farina. Simultaneous measurement of impulse response and distortion with a swept-sine technique. *Presented at the 108th Convention for the audio engineering society 2000 February 19-22 Paris, France, 2000.*
- [10] David Gesbert, Helmut Bolcskei, Dhananjay A. Gore, and Arogyaswami J. Paulraj. Outdoor mimo wireless channels: Models and performance prediction. *IEEE Transaction on Communications*, Vol. 50, NO. 12, DECEMBER, 2002.
- [11] Gene H. Golub and Charles F. Van Loan. *Matrix computations*. The John Hopkins University Press, second edition, 1989.
- [12] Elena Grassi, Jiwanjot Tulsi, and Shihab Shamma. Measurement of head related transfer functions based on the empirical transfer function estimate. *Proceedings of the 2004 International Conference on Auditory Display, Boston, MA, USA, July 6-9, 2003, 2004.*
- [13] Per Christian Hansen. *Rank-Deficient and Discrete Ill-Posed Problems*. SIAM Monographs on Mathematical Modeling and Computation, first edition, 1998.
- [14] William M. Hartmann. How we localize sound. *American Institute of Physics*, <http://www.aip.org/pt/nov99/locsound.html>, 20060623, 1999.
- [15] Monson H. Hayes. *Statistical Digital Signal Processing And Modeling*. John Wiley & Sons, first edition, 1996.
- [16] Markus Hofbauer. *Optimal Linear Separation and Deconvolution of Acoustical Convolutional Mixtures*. PhD thesis, Swiss Federal Institute of Technology, 2005.
- [17] <http://www.oticon.com/>. Harmonisk forvringning. <http://www.oticon.com/eprOtiScripts/Files/encyclopedia/dir.asp?selectedID=440>, 26.07.2006.
- [18] Sabine Van Huffel and Joos Vandewalle. *The Total Least Squares Problem, Computational Aspects And Analysis*. Society for Industrial and Applied Mathematics, first edition, 1991.
- [19] Don H. Johnson and Dan E. Dudgeon. *Array Signal Processing, Concepts And Techniques*. Prentice Hall, first edition, 1993.
- [20] Marcel Joho, Heinz Mathis, and George S. Moschys. An fft-based algorithm for multichannel blind deconvolution. *Circuits and Systems, 1999. ISCAS '99. Proceedings of the 1999 IEEE International Symposium on, Volume: 3, 203-206 vol.3, 1999.*

- [21] Ole Kirkeby, Philip A. Nelson, and Hareo Hamada. Local sound field reproduction using two closely spaced loudspeakers. *The Journal Of The Acoustical Society Of America* VOL. 104(4) October, 1998.
- [22] David C. Lay. *Linear Algebra With Applications*. Addison-Wesley Publishing Company, second edition, 2000.
- [23] Bowon Lee, Camille Goudeseune, and Mark A. Hasegawa-Johnson. Open-loop multichannel inversion of room impulse response. *The Journal of the Acoustical Society of America*, April, Volume 113, Issue 4, pp. 2202-2203, 2003.
- [24] David McGrath. High resolution simulation of acoustic environments. http://www.headwize.com/techlake1_tech.htm, 26.07.2006.
- [25] Nikolas Mitianoudis and Michael E. Davies. Audio source separation of convolutive mixtures. *IEEE Transactions on Speech And Audio Processing*, VOL. 11, NO. 5, September, 2003.
- [26] Masato Miyoshi and Yutaka Kaneda. Inverse filtering of room acoustics. *IEEE Transactions on acoustics, speech, signal processing*, vol. 36, no 2, feb, 1988.
- [27] Swen Müller and Paulo Massarani. Transfer-function measurements with sweeps. *Journal of the Audio Engineering Society*, June, Vol 49 Number 6, 2001.
- [28] Gabriel Montaldo, Geoffroy Lerosey, Arnaud Derode, Arnaud Tourin, Julien de Rosny, and Mathias Fink. Telecommunication in a disordered environment with iterative time reversal. *Waves Random Media* 14, pages 287–302, 2004.
- [29] John N. Mourjopoulos. Digital equalization of room acoustics. *Journal of Audio Engineering Society*, Vol. 42, No. 11, November, 1994.
- [30] Stephen T. Neely and J. B. Allen. Invertibility of a room impulse response. *The Journal of the Acoustical Society of America* VOL. 66(1), pp. 165-169, July, 1979.
- [31] Philip A. Nelson, Felipe Orduna-Bustamante, and Hareo Hamada. Invers filter design and equalization zones in multichannel sound reproduction. *IEEE Transactions on Speech And Audio Processing*, Vol 3, NO.3, May, 1995.
- [32] Proakis and Manolakis. *Digital Signal Processing, Principles, Algorithms, and applications*. Prentice Hall, third edition, 1996.

- [33] Philippe Roux and Mathias Fink. Time reversal in a waveguide: study of the temporal and spatial focusing. *J. Acoust. Soc. Am.* 107 (5), Pt. 1, May, 2000.
- [34] Alastair Sibbald. An introduction to sound and hearing. <http://www.sensaura.co.uk/whitepapers/pdfs/dev005.pdf>, 2001.
- [35] Guy-Bart Stan, Jean Jacques Embrechts, and Dominique Archambeau. Comparison of different impulse response measurement techniques. *The Journal of the Audio Engineering Society*, Vol. 50, No. 4, pp. 249-262, 2002.
- [36] Sylvain Yon, Mickael Tanter, and Mathias Fink. Sound focusing in rooms: The time-reversal approach. *Journal of the Acoustical Society of America* 113(3) March, 2003.
- [37] Sylvain Yon, Mickael Tanter, and Mathias Fink. Sound focusing in rooms.ii. the spatio-temporal inverse filter. *Journal of the Acoustical Society of America* 114(6) Dec, 2003.
- [38] Dmitry N. Zotkin, Ramani Duraiswami, Elena Grassi, and Nail A. Gumerov. Fast head related transfer function measurement via reciprocity. *University of maryland, Computer Science and UMIACS Technical Report, CS-4620 and UMIACS-2004-62*, 2004.

Figurer

| | | |
|------|--|----|
| 1.1 | Forsinkelse-og-sum stråleformer | 6 |
| 1.2 | Oppsett ved tidsreversering | 8 |
| 2.1 | Enkel modell av ekkoer i rom | 15 |
| 2.2 | En impulsrespons målt i et rom | 15 |
| 2.3 | KEMAR-dukken | 18 |
| 2.4 | Tradisjonelt oppsett ved invertering av impulsrespons | 22 |
| 2.5 | Systemoppsett ved invertering av impulsrespons med to høytalere og en mikrofon. | 23 |
| 2.6 | Matrisestruktur for kildeseparasjon med M sensorer og L signalkilder. | 26 |
| 2.7 | Matrisestruktur for lydfokusering med M høytalere og L målepunkter/sensorer. | 27 |
| 2.8 | Hjelpefigur til oppsettet av matrisen \mathbf{H} ved kildeseparasjon. | 28 |
| 2.9 | Systemrepresentasjon av kildeseparasjon med L kilder og M mikrofoner, oppsettet viser separasjon av en kilde. | 29 |
| 2.10 | Hjelpefigur til oppsettet av matrisen \mathbf{H} ved lydfokusering. . . | 30 |
| 2.11 | Systemrepresentasjon av lydfokusering med M høytalere og L mikrofoner. | 31 |
| 3.1 | Reflektorer som vil øke spredningen av lydsignaler. | 37 |
| 3.2 | Reflektorer vil gi opphav til forskjellige impulsresponser i forskjellige punkter rundt lydkildene. | 38 |
| 3.3 | Utsnitt av simulasjonsmodellen ved lydfokusering, ingen elevasjon | 39 |
| 3.4 | Hjelpefigur til sammensetning av den sammensatte konvo- lusjonsmatrisen \mathbf{H} i modellen. | 40 |
| 3.5 | Hjelpefigur for algoritmen som glatter impulsresponsene. . . | 43 |
| 3.6 | To glattingsmønstre | 44 |
| 4.1 | Posisjonene til fokuspunkt og undertrykkelsespunkter tilhø- rende oppsettene (i), (ii), (iii) og (iv) | 49 |
| 4.2 | Posisjonene til fokuspunkt og undertrykkelsespunkter tilhø- rende oppsettene (v), (vi), (vii) og (viii). | 51 |

| | | |
|------|--|----|
| 4.3 | Plott av responsen i fokuspunktet og undertrykkelsespunktene ved forskjellige mengder glatting | 53 |
| 4.4 | Impulsresponser fra kolonnene til de sammensatte konvolusjonsmatrisene \mathbf{H}_1 , \mathbf{H}_2 , \mathbf{H}_3 og \mathbf{H}_4 | 54 |
| 4.5 | Plott av gjennomsnittlige L_2 -verdier fra modeller med 2 lydkilder, 2 undertrykkelsespunkter, og ett fokuspunkt. Glattingsmønster, 1(a) og 2(b), algoritmen utført henholdsvis 3(a) og 6(b) ganger. Med de tilhørende gjennomsnittlige utilsiktede relative energinivåene. | 56 |
| 4.6 | Plott av gjennomsnittlige L_2 -verdier fra modeller med 4 lydkilder, 4 undertrykkelsespunkter og ett fokuspunkt. Glattingsmønster, 1(a) og 2(b), algoritmen utført henholdsvis 3(a) og 6(b) ganger. Med de tilhørende gjennomsnittlige utilsiktede relative energinivåene. | 58 |
| 4.7 | Fordeling av relative energinivåer ved forskjellige impulsresponslengder. Den øverste fordelingen er fra modeller med impulsresponslengde 10, den nest øverste fra modeller med impulsresponslengde 100, den tredje fra modeller med impulsresponslengde 200 og den siste fra modeller med impulsresponslengde 500. | 60 |
| 4.8 | Fordeling av relative maksimale utslag for forskjellige impulsresponslengder. Den øverste fordelingen er fra modeller med impulsresponslengde 10, den nest øverste fra modeller med impulsresponslengde 100, den tredje er fra modeller med impulsresponslengde 200 og den siste fra modeller med impulsresponslengde 500. | 61 |
| 4.9 | Fordeling av tilpasningsfeil ved forskjellige impulsresponslengder. Den øverste fordelingen er fra modeller med impulsresponslengde 10, den nest øverste fra modeller med impulsresponslengde 100, den tredje er fra modeller med impulsresponslengde 200 og den siste fra modeller med impulsresponslengde 500. | 61 |
| 4.10 | 40 plott av et utvalg av L_2 -verdier fra modeller med impulsresponser av lengde 10(a), 100(b), 200(c) og 500(d). De røde plottene er gjennomsnittene av de 40 realiseringene. | 62 |
| 4.11 | 40 plott av et utvalg av L_∞ -verdier fra modeller med impulsresponser av lengde 10(a), 100(b), 200(c) og 500(d). De røde plottene er gjennomsnittene av de 40 realiseringene. | 63 |
| 4.12 | Plott av L_2 -verdiene ved enkelt realiseringer av modellen. Modellene har 4 lydkilder, ett fokuspunkt og 4 undertrykkelsespunkter og er påført glattingsskjema 1. | 64 |
| 4.13 | Plott av L_2 -verdiene ved enkelt realiseringer av modellen. Modellene har 4 lydkilder, ett fokuspunkt og 4 undertrykkelsespunkter og er påført glattingsskjema 1. | 64 |

| | | |
|------|--|----|
| 4.14 | Plott av L_2 -verdiene ved enkelt realiseringer av modellen. Modellene har 10 lydkilder, ett fokuspunkt og 10 undertrykkelsespunkter og er påført glattingsskjema 1. | 65 |
| 4.15 | Plott av L_2 -verdiene ved enkelt realiseringer av modellen. Modellene har 10 lydkilder, ett fokuspunkt og 10 undertrykkelsespunkter og er påført glattingsskjema 1 | 65 |
| 4.16 | Gjennomsnittlige L_2 -verdier plottet fra modeller med 2 lydkilder og forskjellige antall undertrykkelsespunkter. . . . | 67 |
| 4.17 | Gjennomsnittlige L_2 -verdier plottet fra modeller med 4 lydkilder og forskjellige antall undertrykkelsespunkter. . . . | 68 |
| 4.18 | Plassering av kontrollpunktene i modell-oppsettene 1,2,3 og 4. | 70 |
| 4.19 | Plott av gjennomsnittlige L_2 -verdier fra modeller med modell-oppsett 1, 2, 3, og 4, med tilhørende gjennomsnittlige verdier på utilsiktet undertrykking. | 70 |
| 4.20 | Plassering av kontrollpunktene i modell-oppsettene 5 og 6. . | 71 |
| 4.21 | Plott av gjennomsnittlige L_2 -verdier fra modeller med modell-oppsett 5 og 6, med tilhørende gjennomsnittlige utilsiktede undertrykkingsverdier. | 72 |
| 5.1 | Bilde av oppsettet | 74 |
| 5.2 | Oppsettet av systemet ved målinger av impulsresponser i rom. | 75 |
| 5.3 | Hjelpefigur til sammensetning av den sammensatte konvolusjonsmatrisen | 76 |
| 5.4 | Plott av fokuseringen med impulsresponser med og uten reflektorer, L_2 verdiene plottet | 77 |
| 5.5 | Plott av fokuseringen med impulsresponser med og uten reflektorer, L_∞ verdiene plottet | 78 |
| 5.6 | L_2 -verdiene plottet hvor det er lagt støy på impulsresponsene | 81 |
| 5.7 | Hjelpefigur til sammensetningen av matrisen \mathbf{H} ved kilde-separasjon med HRIR. | 83 |
| 5.8 | “Hode-plott” med fokus inn på en kilde i punktet (85,0). Sylindren merket A, skjærer kulen der den ønskede kilden er plassert. Sylindrene merket B og C skjærer kulen der de uønskede kildene er plassert. | 84 |
| 5.9 | Overblikk av separasjonen med hode-relaterte impulsresponser, ved kildeplassering 1. | 84 |
| 5.10 | “Hodeplott” med fokus inn på en kilde plassert i punktet (-45,65). Sylindren merket A, skjærer kulen der den ønskede kilden er. Sylindrene merket B og C skjærer kulen der de uønskede kildene er. | 85 |
| 5.11 | Overblikk av separasjonen med hode-relaterte impulsresponser ved kildeplassering 2. | 86 |
| 5.12 | Impulsresponser som inngår i den sammensatte konvolusjonsmatrisen \mathbf{H} ved kildeplassering 1. | 88 |

| | | |
|------|--|----|
| 5.13 | “Hodeplott” med fokus inn på en kilde rett foran lytteren. Sylinderen merket A, skjærer kulen der den ønskede kilden er. Sylinderne merket B og C skjærer kulen der de uønskede kildene er. | 88 |
| A.1 | Sinussignal med eksponensielt økende frekvens, 0.5 sekun- der av et målesignal på 1 sekund | 96 |

Tabeller

| | | |
|-----|---|----|
| 4.1 | Gjennomsnittlige relative energinivåer, mellom fokuspunkt og undertrykkelsespunkter, fra modeller med to lydkilder, ett fokuspunkt, to undertrykkelsespunkter og forskjellig mengder glatting | 50 |
| 4.2 | Gjennomsnittlige relative energinivåer, mellom fokuspunkt og undertrykkelsespunkter, fra modeller med 4 lydkilder, ett fokuspunkt, 4 undertrykkelsespunkter og forskjellig mengder glatting | 52 |
| 4.3 | Gjennomsnittlige relative energinivåer, relative maksimale utslag og tilpasningsfeil fra modeller med impulsresponser av lengde 10, 100, 200 og 500 | 59 |
| 4.4 | Gjennomsnittlige relative energinivåer, utilsiktet undertrykking og tilpasningsfeil for modeller med 2 lydkilder og med forskjellige antall undertrykkelsespunkter | 66 |
| 4.5 | Gjennomsnittlige relative energinivåer, utilsiktet undertrykking og tilpasningsfeil for modeller med 4 lydkilder med forskjellige antall undertrykkelsespunkter | 67 |
| 4.6 | Gjennomsnittlige relative energinivåer, relative maksimale utslag og tilpasningsfeil for simuleringer hvor de inverse filterene er lengre enn impulsresponsene. | 68 |
| 5.1 | Tilpasningsfeilen og relative energinivåer ved fokusering under støyete forhold | 81 |